Proceedings of the 7th
World Congress on Intelligent Control and Automation
June 25 - 27, 2008, Chongqing, China

An Importance Sampling Method Based on Martingale with Applications to Rare Event Probability*

Yue Qiu and Hong Zhou
School of Economics and Management
Beihang University
Beijing 100083, P.R China
qiuyueqq@sem.buaa.edu.cn
h_zhou@buaa.edu.cn

Yueqin Wu
School of Economics and Management
Beihang University
Beijing 100083, P.R China
wuyueqin01@sina.com

***Abstract* - It usually takes long time to simulate rare event using traditional Monte Carlo method, while importance sampling techniques can effectively reduce the simulation time and improve simulation efficiency. A new implementation for importance sampling method to estimate rare event probability in simulation models is proposed. The optimal importance sampling distributions was obtained by making use of the martingale constructed by likelihood ratio. The computation results were compared with the importance sampling based on cross-entropy, the importance sampling based on minimizing variance and crude Monte Carlo method. Numerical experiments had been conducted and the results indicate that the method can effectively estimate the rare event probabilities.**

Index Terms* - *rare event; importance sampling; martingale; likelihood ratio

I. INTRODUCTION

Rare event simulation has attracted extensive attentions since the concept was first proposed, but there haven't been many research results yet [7], [8], [23]. Performances of modern systems, such as coherent reliability systems, inventory systems, insurance risk, storage systems, computer networks and telecommunications networks, are sometimes characterized by probabilities of rare events and usually studied by simulation [1],[2],[5],[9],[12],[16],[17],[19],[25]. Straightforward simulation for rare events requires a large number of trials and hard to implement because the occurrence of rare events are extremely little in a standard simulation [6], [10], [11], [13], hence new methods are needed to be investigated and developed. One way to overcome this difficulty is to use importance sampling (IS). The main idea of IS is to make the occurrence of rare events more frequent by carrying out the simulation with a different probability distribution – what is the so-called change of measure (CM) – and to estimate the probability of interest via a corresponding likelihood ratio (LR) estimator[3],[14].

It is commonly known that the greatest difficulty in obtaining an efficient IS method is to find out the optimal importance sampling distribution function. It has been shown that, in many cases, the computation and experiments required for achieving an efficient IS method may be even more complicated than the original problem. Thus, this implies that we have to find simple and efficient IS algorithms. There are two main ways to solve this problem. One is to utilize Variance Minimization (VM) technique [14],[18]; the other is to use minimizing Cross Entropy (CE) technique[4],[14],[15]. It is well known that there theoretically exists a CM that yields a zero-variance likelihood ratio estimator. However, in practice such an optimal CM cannot be computed since it depends on the underlying quantity being estimated. However, investigations have shown that the reduced variance method can be employed to approach the optimal solution in the sense of minimizing the variance [4], [20], [21]. In other words, minimizing estimator's variance with original density function is equivalent to minimizing the expected likelihood ratio conditioned on the rare event happening; and minimizing cross entropy is equivalent to minimizing the expected logarithm of likelihood ratio conditioned on the rare event happening. Therefore, minimizing cross entropy in a sense is close to, but definitely is different from minimizing estimator's variance. In terms of estimator's variance, cross entropy method does not seek for the optimal solution [22].

To reduce the number of simulation and improve the efficiency of sampling, a novel method of estimating the probability of the rare event is presented in this paper. The optimal importance sampling distributions was obtained by making use of the martingale constructed by likelihood ratio. We estimate the rare event probability via minimize the distance between expectation of likelihood ratio and 1. Moreover, the importance sampling distribution is chosen from the same parametric family as the transformed distribution. And the optimal parameter vector of the importance sampling distribution is estimated by minimizing the distance between the optimal importance sampling distribution $g_{opt}(x)$ and $f(\cdot;v)$ .The computation results were compared with the importance sampling based on cross-entropy, the importance sampling based on minimizing variance and crude Monte Carlo method. Numerical experiments had been conducted and the results indicate that the method can effectively estimate the rare event probabilities. The obtained numerical results suggest that the method can effectively estimates rare event probability.

II. IMPORTANCE SAMPLING

* This work is support from the National Natural Science Foundation of China (Project no. 70531010, 70771003), and Program for New Century Excellent Talents in University of China (Project no. NCET-04-0175).

IS is based on the idea to make the occurrence of rare events more frequent, or in other words, to speed up the simulation. Technically, IS aims to select a probability distribution that minimizes the variance of the IS estimate. The efficiency of IS depends on obtaining a good change of measure. Large deviation results have proven their applicability for small systems with medium complexity [8], [14], [23].

Let $x=(X_1,\ldots,X_n)$ be a random vector taking values in some space Ω and $f(x)$ be a probability density function on Ω. Let s be some real-valued function on Ω. Suppose we are interested in the probability that $s(x)$ is greater than or equal to some real number r, this probability can be expressed as $l=P\{s(x)\geq r\}=EI_{(s(x)\geq r)}$. If this probability is very small, say smaller than 10^{-5}, we call $\{s(x)\geq r\}$ a rare event.

A straightforward way to estimate $l=P\{s(x)\geq r\}=EI_{(s(x)\geq r)}$ is to use crude Monte-Carlo simulation: Draw a random sample $X_1,\ldots,X_N$ from $f(x)$; then

$$\begin{aligned}\hat{l}&=\frac{1}{N}\sum_{i=1}^{N}I_{\{s(X_i)\}}\\&=\frac{1}{N}\sum_{i=1}^{N}I_{\{s(X_i)\geq r\}}\end{aligned}. \tag{1}$$

is an unbiased estimator of l.

Where,

$$I_{\{s(X_i)\geq r\}}=\begin{cases}1 & s(X_i)\geq r\\ 0 & s(X_i)<r\end{cases} \tag{2}$$

However this poses serious problems when $\{s(x)\geq r\}$ is a rare event. In that case a large simulation effort is required in order to estimate l accurately, i.e., with small relative error or a narrow confidence interval.

An alternative is based on importance sampling: take a random sample $X_1,\ldots,X_N$ from an importance sampling density $g(x)$ on Ω, then

$$\begin{aligned}l&=\int_{\{s(x)\geq r\}}f(x)dx\\&=\int_{\{s(x)\geq r\}}L(x)g(x)dx\end{aligned}. \tag{3}$$

where $L(x)=\dfrac{f(x)}{g(x)}$ is called the likelihood ratio. It is well known that the best way to estimate l is to use the change of measure with density

$$g_{opt}(x)=\frac{I_{(s(x)\geq r)}f(x)}{r}. \tag{4}$$

The estimator (3) has zero variance, and we need to produce only N = 1 sample. The obvious difficulty is of course that this $g_{opt}(x)$ depends on the unknown parameter, so it cannot be computed in practice. In estimate rare event probability: Draw a random sample $X_1,\ldots,X_N$ from $g(x)$, and then get estimator according to (2)

$$\hat{l}_{IS}=\sum_{i=1}^{R}I_{\{s(X_i)\geq r\}}L(X_i) \tag{5}$$

We will call the estimator in (4) the *importance sampling estimator.*

To obtain the optimal – or at least a good – change of measure from the original sampling distribution $f(x)$ to a new distribution $g(x)$, is the main challenge with respect to making importance sampling efficient and robust. From equation (3), the only restriction to the probability density $g(x)$ is that $g(x)>0$ for all samples x. This is a necessary condition which serves as a guideline for choosing $g(x)$. The efficiency of the importance sampling estimate in (5) is dependent on an appropriate choice of the new distribution. An unfortunate choice of $g(x)$ may cause the variance of (5) to be larger than the variance of the estimates obtained by direct simulation. For some $g(x)$, the variance may be infinite, and, hence, it is crucial to find a good $g(x)$.

III. Importance Sampling method Based on the Martingale

As discuses above, the obvious difficulty is of course that this $g_{opt}(x)$ depends on the unknown parameter l. Moreover, it is often convenient to choose a $g(x)$ in the family of densities $\{f(.;v)\}$. The idea now is to choose the parameter vector, called the reference parameter v such that the distance between the densities $g_{opt}(x)$ and $f(.;v)$ is minimal.

Martingale: A process $\{X_n,n\geq 0\}$ is called a martingale, if $\forall n\geq 0$, we have

(1) $E|X_n|<\infty$,

(2) $E(X_{n+1}|X_0,X_1,\cdots X_n)=X_n \quad a.s.$

Sometimes $\{X_n,n\geq 0\}$ can not be straight observed, so we observe another process $\{Y_n,n\geq 0\}$. We have:

Let $\{X_n,n\geq 0\}$ and $\{Y_n,n\geq 0\}$ are two processes, $\{X_n,n\geq 0\}$ is called a martingale about $\{Y_n,n\geq 0\}$, if we have

(1) $E|X_n|<\infty$,

(2) $E(X_{n+1}|Y_0,Y_1,\cdots Y_n)=X_n \quad a.s.$

Martingale constructed by likelihood ratio：Let $Y_0, Y_1, \ldots, Y_n, \ldots$ are independent random variables with identical distribution, f_0 and f_1 are probability density functions，Let

$$X_n = \frac{f_1(Y_0) f_1(Y_1) \cdots f_1(Y_n)}{f_0(Y_0) f_0(Y_1) \cdots f_0(Y_n)}, \quad n \ge 0$$

Suppose $f_0(y) > 0$ for $\forall y$, When the probability density function of Y_n is f_0, then $\{X_n, n \ge 0\}$ is called a martingale about $\{Y_n, n \ge 0\}$ .It is common to use $\{X_n, n \ge 0\}$ is a martingale about $\{Y_n, n \ge 0\}$ to ensure the probability density function of $\{Y_n, n \ge 0\}$ is f_0.

For

$$\begin{aligned} &E|X_n| \\ &= E\left[\frac{f_1(Y_0) f_1(Y_1) \cdots f_1(Y_n)}{f_0(Y_0) f_0(Y_1) \cdots f_0(Y_n)}\right] \\ &= 1 < \infty \end{aligned}$$

and

$$\begin{aligned} &E(X_{n+1}|Y_0, Y_1, \cdots Y_n) \\ &= E\left[X_n \frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}|Y_0, Y_1, \cdots Y_n\right] \\ &= X_n E\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right] \end{aligned}$$

When the probability density function of Y_n is f_0, then

$$E\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right] = \int \frac{f_1(y)}{f_0(y)} f_0(y) dy = \int f_1(y) dy = 1$$

$\{X_n, n \ge 0\}$ is a martingale about $\{Y_n, n \ge 0\}$.

As discuses above, $\{X_n, n \ge 0\}$ is a martingale about $\{Y_n, n \ge 0\}$ when the probability density functions of $\{Y_n, n \ge 0\}$ is f_0 . Minimizing the distance between $g_{opt}(x)$ and $f(.;v)$ is equivalent to choosing f_0 as the optimal importance sampling distribution function $g_{opt}(x)$ ，$\{X_n, n \ge 0\}$ is called a martingale about $\{Y_n, n \ge 0\}$ when generate Y_{n+1} from density function $f(.;v)$ in some special situation. Which is equivalent to choosing v by minimize $\left|E_v\left[\frac{f(Y_{n+1} \cdot ;U)}{g_{opt}(Y_{n+1})}\right] - 1\right|$.

Substituting $g_{opt}(x)$ from (4), we obtain the minimization program

$$\min_v \left| E_v \frac{l}{I_{\{s(x) \ge r\}}} - 1 \right| \qquad (6)$$

Using again importance sampling, we can rewrite (6) as

$$\min_v \left| E_w \frac{l}{I_{\{s(x) \ge r\}}} W(X; v, w) - 1 \right| \qquad (7)$$

where，$W(x; v, w) = \dfrac{f(x; v)}{f(x; w)}$ is the likelihood ratio.

It is important to note that the algorithm is useful only in the case where the probability of the event $\{s(X) \ge r\}$ is not too small. However, due to the rareness of the events, most of the indicator random variables $I_{\{s(X_i) \ge r\}}, i = 1, \ldots, N$ will be zero, for moderate N . A multi-level algorithm can be used to overcome this difficulty. The idea is to construct a sequence of reference parameters $\{v_t, t \ge 0\}$ and a sequence of levels $\{r_t, t \ge 1\}$,and iterate in both r_t and v_t ,such that $p_{v_{t-1}}(s(x) \ge r_t) \ge \rho$ ($\rho \in (0,1)$ is defined in advance）.

Remark 3.1

The indicator random variables $I_{\{s(X_i) \ge r\}}, i = 1, \ldots, N$ in (7) is in the position of denominator, so we redefine the indicator function of (2) as

$$I_{\{s(X_i) \ge r\}} = \begin{cases} 1, & s(X_i) \ge r, \\ 10^{-5}, & s(X_i) < r. \end{cases} \qquad (8)$$

We use iterative method to overcome the difficulty that l in (7) is unknown parameter. Construct a sequence of reference parameters $\{l_t, t > 0\}$ to solve this problem, i.e.

$$l_t = \frac{1}{N} \sum_{i=1}^{N} I_{\{s(X_i) \ge r_t\}} W(X_i; u, v_{t-1}) \qquad (9)$$

The above rationale results in the following algorithm:

Step1. Define $v_0 = u$, $\rho = 0.1$;

Step2. Generate a sample $X_1, \ldots, X_N$ from importance sampling probability density function $f(.; v_{t-1})$, and compute the sample $(1-\rho)$-quantile r_t of the performances, provided $r_t < r$,otherwise set $r_t = r$ .update l_t according (9);

Step3. Use the same sample $X_1, \ldots, X_N$ in *Step 2* to solve the stochastic program (7). Denote the solution by v_t .

Step4. If $r_t < r$, set $t = t+1$ and reiterate from *Step 2,* Else proceed with *Step 5;*

Step5. Estimate the rare-event probability using the importance sampling estimator (5).

Flow chart of algorithm is followed：

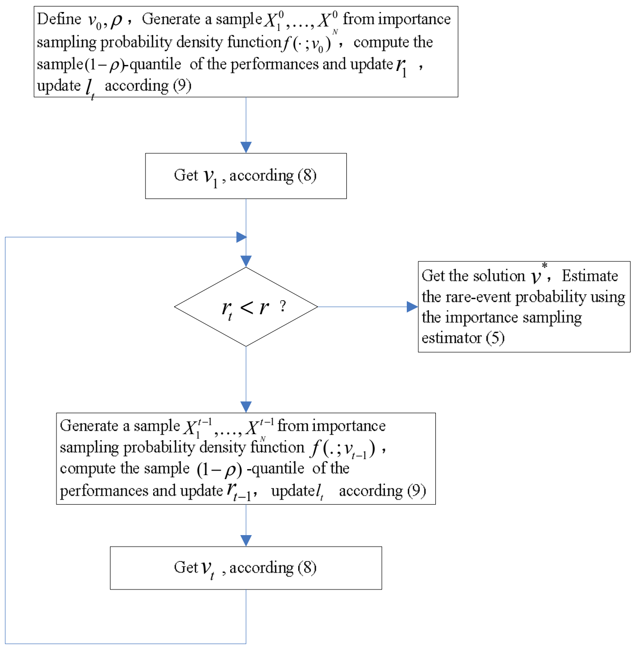

Fig. 1 Flow chart of important sampling algorithm based on the martingale constructed by likelihood ratio

IV. NUMERICAL EXAMPLES

To illustrate the ideas set forth in the previous sections, we present now numerical results obtained for project management problem. There are 5 key activities on key routes in a project. Let x denote the man-hour of all activities, and let $s(x)$ denote the total completion time of this project. Our goal is to estimate the probability that the total completion time of this project will be more than a given time r. That is, we want to estimate $l = P(s(x) > r)$. If $l < 10^{-5}$, and this is a rare event. We consider the case where all man-hours of all activities have exponential distribution. For simplicity, we assume that the times of all activities are i.i.d. with mean $\mu = 25$. The above procedure was replicated 100 times, and the average and a simultaneous 90% confidence interval were built from those 100 independent estimates, both for the two methods. We consider two values for r, namely, $r = 500$ $r = 700$. The following tables displays the estimation results. In the table, $\hat{l}$ is the estimate for $l = P(s(x) > r)$, "90% H.W." denotes the half-width of a 90% confidence interval and N is the sample size.

Table 2 presents the performance of the importance sampling method based on martingale for the problem of estimating rare event probability $l = P(s(x) > r)$. To show the efficiency of the method, we used the Monte Carlo method, CE method and the importance sampling based on Variance Minimization to estimate $l = P(s(x) > r)$ at the same time.

TABLE I
Simulation result based on Monte Carlo method

r	$\hat{l}$	90%H.W.	N
500	1.0000e-006	3.2800e-006	200000
700	—	—	5000000

TABLE II
Simulation result based on martingale method

r	$\hat{l}_{IS}$	90%H.W.	N
500	1.6177e-006	6.5440e-006	1000
700	3.0627e-011	2.3219e-011	2000

TABLE III
Simulation result based on CE method

r	$\hat{l}_{IS}$	90%H.W.	N
500	2.1804e-006	2.6408e-006	1000
700	2.7474e-011	2.7874e-011	2000

TABLE IV
Simulation result based on Variance Minimization method

r	$\hat{l}_{IS}$	90%H.W.	N
500	9.3614e-006	1.3451e-005	1000
700	4.6326e-011	6.7432e-011	50000

Note: "—"in table 1 denote rare event did not happen after 5000000 samples and could not establish effective interval estimation.

It is obviously to see from table I and table II that the high efficiency of the martingale method for estimating rare events probabilities, where the naive Monte Carlo method fails. For events that are not very rare, the martingale method may still help in terms of providing estimates with smaller variance. Suppose we wish to estimate the probability that the total completion time of this project is greater than $r = 500$ .The crude Monte Carlo method with $N = 200000$ gave an estimate 1.0000e-006 with an estimated half width of confidence interval of 3.2800e-006; using the estimated optimal parameter vector, the final step with $N = 1000$ gave an estimate of 1.6177e-006 with an estimate half width of confidence interval of 6.5440e-006, which reduced our simulation effort quickly. Suppose we wish to estimate the probability that the total completion time of this project is greater than $r = 700$ .Rare event did not happen after $N = 5000000$ samples and could not establish effective interval estimation use Monte Carlo method. However, the algorithm proposed in this paper gets satisfying result with $N = 2000$.

Suppose we wish to estimate the probability that the total completion time of this project is greater than $r = 500$ .It is convenient to see from comparison between table II and table III, table IV that the estimate precision of martingale is worse than CE method, but better than the variance minimization method. Suppose we wish to estimate the probability that the total completion time of this project is greater than $r = 700$. The CE method gave an estimate 2.7474e-011 with an estimated half width of confidence interval of 2.7874e-011; the variance minimization method gave an estimate 4.6326e-011 with an estimated half width of confidence interval of 6.7432e-011; while the martingale can gave an estimate 3.0627e-011 with an estimated half width of confidence interval of 2.3219e-011.From the discussion above, the martingale method is more reliability for estimate rare event probability.

V. CONCLUSION

In this paper, an important sampling method based on martingale is presented. The key concept is to seek the importance sampling distribution function by ensuring the $\{X_n, n \geq 0\}$ constructed by likelihood ration is a martingale about $\{Y_n, n \geq 0\}$ when the probability density function of independent random variables with identical distribution $\{Y_n, n \geq 0\}$ is the optimal importance sampling distribution function. Then generates samples from importance sampling distribution function, get the probability of the rare event at the end. We present simulation result which supports the efficiency of the method. Rare event simulation research is still at the primary stage, there is a broad space for development both in theory and in application [4], [13], [14], [23]. Currently, we are extending this method for other types of applications, such as economic crisis and the early-warning of financial risk.

ACKNOWLEDGMENT

The authors would express sincere appreciation to the support from the National Natural Science Foundation of China (Project no. 70531010, 70771003), and Program for New Century Excellent Talents in University of China (Project no. NCET-04-0175).

REFERENCES

[1] Reuven Y. Rubinstein. The Cross-Entropy Method for Combinatorial and Continuous Optimization [J]. Methodology and Computing in Applied Probability, 1999, 1: 127-190

[2] Alon, G., D.P. Kroese, T. Raviv, and R.Y. Rubinstein.. Application of the Cross-Entropy Method to the Buffer Allocation Problem in a Simulation-Based Environment.[J] Annals of Operations Research 134(2005), 19–67.

[3] D.Siegmund. Importance Sampling in the Monte Carlo Study of Sequential Tests. The Annals of Statistics. 4(1976) 673-684

[4] Andre Costaa, Owen Dafydd Jones, Dirk Kroese, Convergence properties of the cross-entropy method for discrete optimization[J], Operations Research Letters 35 (2007) 573 – 580

[5] Paul Glasserman, Jingyi Li Importance Sampling for Portfolio Credit Risk[J Management Science 2005, V51, N11,1643 – 1656

[6] Tito Homem-de-Mello and R.Y.Rubinstein, “Estimation of Rare Event Probabilities Using Cross-Entropy,” Proceedings of the 2002 Winter simulation Conference,2002,pp. 310-319.

[7] Hammersley, J, M, and D. C, Handscomb. Moute Carlo Methods[M], Methuen, London, 1964.

[8] M Hsieh. Adaptive Importance Sampling for Rare Event Simulation of Queuing Networks [D]. Ph. D. thesis, Stanford University. California,1997

[9] Hong ZHOU,Yue QIU, Yue-qin WU. An early warning system for loan risk assessment based on rare event simulation[J]. Asia Simulation Conference 2007. Accepted

[10]Kroese, D.P. and R.Y. Rubinstein. The Transform Likelihood Ratio Method for Rare Event Simulation with Heavy Tails[J]. Queueing Systems,2004, 46, 317–351.

[11]Michael Edward Larson. Stochastic Optimization of Rare Event Probability Problems. PhD Thesis, Harvard University, the American, 1996

[12]Cohen, I., B. Golany, and A. Shtub. Managing Stochastic Finite Capacity Multi-Project Systems Through the Cross-Entropy Method[J]. Annals of Operations Research,2005 134, 183–199.

[13]JA Bucklew, Introduction to Rare Event Simulation [M] Springer,New york 2004

[14]Pieter-Tjerk De Boer , D.P. Kroese , Shie Mannor. A Tutorial on the Cross-Entropy Method [J]. Annals of Operations Research, 2005,134: 19–67

[15]R.Y. Rubinstein. A Stochastic Minimum Cross-Entropy Method for Combinatorial Optimization and Rare-Event Estimation [J]. Methodology and Computing in Applied Probability, 2005, 7: 5–50

[16]Al-Qaq WA, Devetsiklotis M, Townsend J K. Stochastic Gradient Optimization of Importance Sampling for the Efficient Simulation of Digital Communication Systems[J]. IEEE Trans. Communications. 43(1995),2975-2985

[17] Chepuri, K. and T. Homem-de-Mello. “Solving the Vehicle Routing Problem with Stochastic Demands using the Cross-Entropy Method.” Annals of Operations Research 134(2005)., 53–181.

[18] Devetsikiotis M, Townsend J K. an Algorithmic Approach to the Optimization of Importance Sampling Parameters in Digital Communication System Simulation. IEEE Trans. Communications.41(1993)1464-1473

[19] Hui, K.-P., N. Bean, M. Kraetzl, and D.P. Kroese. “The Cross-Entropy Method for Network Reliability Estimation.” Annals of Operations Research 134(2005)., 101–118.

[20] Xingbin Yu, Chanan Singh, “Expected Power Loss Calculation Including Protection Failures using Importance Sampling and SOM,” *Power Engineering Society General Meeting,* 2004. IEEE, Vol.1, pp. 206- 211

[21] Qiuming Zhu, “Minimum Cross-Entropy Approximation for Modeling of Highly Intertwining Data Sets at Subclass Levels,” *Journal of Intelligent Information Systems,* 1998, vol 11pp. 139–152

[22]M Villén-Altamirano, A Martínez-Marrón, J Gamo, F Fernández- Cuesta. Enhancement of the Accelerated Simulation Method RESTART by considering Multiple Thresholds[C]// Proceedings 14th International Tele traffic Congress.,The fundaaniental role of teletraffic in the evolution of telecomririunications networks, Ed J. Labetoulle, and J.W. Roberts, Elsevier.1994: 797–810.

[23] R.Y. Rubinstein, D.P. Kroese. The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning[M]. Springer-Verlag, 2004.

[24]Reuven Y. Rubinstein. Optimization of computer simulation models with rare events[J]. European Journal of Operational Research 99 (1997) 89-112

[25]Pieter-Tjerk De Boer. Rare-Event Simulation of Non-Markovian Queueing Networks Using a State-Dependent Change of Measure[J]. Annals of Operations Research 2005, 134: 69–100