



## Management Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### A Partitioning Algorithm for Markov Decision Processes with Applications to Market Microstructure

Ningyuan Chen, Steven Kou, Chun Wang

To cite this article:

Ningyuan Chen, Steven Kou, Chun Wang (2017) A Partitioning Algorithm for Markov Decision Processes with Applications to Market Microstructure. Management Science

Published online in Articles in Advance 17 Feb 2017

<https://doi.org/10.1287/mnsc.2016.2639>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# A Partitioning Algorithm for Markov Decision Processes with Applications to Market Microstructure

Ningyuan Chen,<sup>a</sup> Steven Kou,<sup>b</sup> Chun Wang<sup>c</sup>

<sup>a</sup>Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong; <sup>b</sup>Risk Management Institute, National University of Singapore, Singapore 119077; <sup>c</sup>Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027

Contact: [nychen@ust.hk](mailto:nychen@ust.hk) (NC); [matsteve@nus.edu.sg](mailto:matsteve@nus.edu.sg) (SK); [cw2519@columbia.edu](mailto:cw2519@columbia.edu) (CW)

Received: January 20, 2015

Accepted: August 12, 2016

Published Online in Articles in Advance:  
February 17, 2017

<https://doi.org/10.1287/mnsc.2016.2639>

Copyright: © 2017 INFORMS

**Abstract.** We propose a partitioning algorithm to solve a class of linear-quadratic Markov decision processes with inequality constraints and nonconvex stagewise cost; within each region of the partitioned state space, the value function and the optimal policy have analytical quadratic and linear forms, respectively. Compared to grid-based numerical schemes, the partitioning algorithm gives the closed-form solution without discretization error, and in many cases does not suffer from the curse of dimensionality. The algorithm is applied to two applications. In the main application, we present a model for limit order books with stochastic market depth to study the optimal order execution problem; stochastic market depth is consistent with empirical studies and necessary to accommodate various order activities. The optimal execution policy obtained by the algorithm significantly outperforms that of a deterministic market depth model in numerical examples. In the second application, we use the algorithm to compute the exact optimal solution to the renewable electricity management problem, for which previously only an approximate solution was known. As a comparison, we show that the approximate solution can be quite inaccurate for some initial states and thus demonstrate an advantage of the exact solution.

**History:** Accepted by Yinyu-Ye, optimization.

**Funding:** The research of S. Kou was supported in part by the Ministry of Education in Singapore [Grant MOE2014-T2-1-006] and by the National Natural Science Foundation of China [Grant R-2015-S-005]. The research of N. Chen was supported in part by the Hong Kong University of Science and Technology [Start-up Fund R9382].

**Supplemental Material:** The online appendix is available at <https://doi.org/10.1287/mnsc.2016.2639>.

**Keywords:** Markov chains • large order execution • electricity trading/production • partitioning • quadratic stochastic programming

## 1. Introduction

Linear-quadratic-Gaussian control is one of the few examples in stochastic dynamic programming that yields a closed-form solution (see, e.g., Bertsekas 1995, Athans 1971). It has quadratic costs and linear state dynamics; the optimal policy is a linear function, and the optimal value function is a quadratic function of the state variables. However, two limitations may exist: First, if there are constraints on actions (decision variables), even if they are linear, obtaining analytical solutions may be difficult. Second, the noise in the real world may be non-Gaussian and intertemporally correlated.

To address these problems, we study a class of optimization problems with linear dynamics, quadratic costs, and linear inequality constraints on the action. Furthermore, the uncertainty is modeled as an underlying Markov chain, which fits into the framework of Markov decision processes (MDPs) (see, e.g., Puterman 2009). More precisely, the coefficients, cost matrix, and constraints may jump, contingent on the Markov chain, and the form of dependence is arbitrary (i.e., not necessarily linear or quadratic). In addition, we relax

the convexity assumption (e.g., in Birge and Louveaux 2011 and Chizeck et al. 1986) that the stagewise cost function must be positive definite. This relaxation is crucial to accommodate the two applications we are interested in exploring.

Rather than using approximations, we propose a partitioning algorithm to characterize the analytical solution of this class of MDPs. Indeed, within each region of the partitioned state space, the value function and the optimal policy have analytical quadratic and linear forms, respectively. Hence, our algorithm is exempt from discretization error, and the complexity usually does not exponentially scale in the dimension of the state space, which are two major drawbacks of grid-based numerical methods.

We apply the algorithm to two applications related to market microstructure. For the first application of the optimal order execution problem in a limit order book (LOB), the general class of MDPs allows us to introduce stochastic market depth, a desirable property of LOB models. The resulting (analytically solved) optimal execution policy is stochastic, and it significantly outperforms deterministic order execution policies in

our numerical examples. For the second application of renewable electricity management in intraday markets, the partitioning algorithm can analytically solve the discrete-time version of the optimal power trading/production problem in Aid et al. (2016), which has, to our knowledge, previously been solved only approximately. It is numerically demonstrated that for some initial states the approximate solution can perform rather poorly.

### 1.1. Background on Market Microstructure

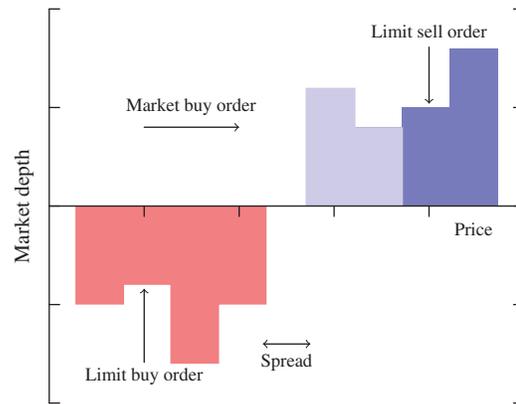
Market microstructure concerns how a transaction occurs and its impact on future prices. For example, in LOBs, the execution price essentially depends on the market depth of the LOB. In the model of renewable electricity management, trading activities in intraday markets lead to price impacts that may affect future transactions. Next, we briefly mention the backgrounds of the two applications.

An LOB is the list of limit orders that a trading venue uses to record the interest of buyers and sellers. Each limit order specifies a quantity, the intention to buy or sell, and a reservation price. For example, a limit buy order may be placed to purchase 10 shares at \$20. Clearly, no limit sell order sets an ask price lower than the bid price of any limit buy order; otherwise, a transaction would occur. Hence, the highest bid price and the lowest ask price give rise to the so-called bid-ask spread. All limit orders in the LOB constitute a profile that reflects demand/supply at each price. Thus, the LOB is characterized by its *market depth*, i.e., the density of limit order shares at each price. In addition to limit orders, there are market orders, which specify only the intention to buy or sell and a quantity for immediate execution. For example, if a market buy order to purchase 10 shares is submitted, then it is immediately matched to 10 shares of limit sell orders with the lowest ask prices, and is executed at the prices specified by those limit orders. A first-come-first-serve rule is triggered if multiple limit orders are placed on the same price. A limit order book is illustrated in Figure 1.

An important question related to LOB models is the optimal large order execution commonly faced by institutional traders, who may submit large orders with significant market impacts (see, e.g., Saar 2001, Chiyachantana et al. 2004). For example, a trader may want to buy 1,000,000 shares in a day. How should the order be split into small market orders over time so that the execution cost is minimized? Our paper addresses this problem with a new LOB model in which the market depth is stochastic and is governed by a Markov chain.

We also study the renewable electricity management problem discussed in Aid et al. (2016) as another application of the partitioning algorithm. According to Henriot (2014), it is important for renewable energy producers to use intraday markets to balance

Figure 1. (Color online) Illustration of a Limit Order Book



Notes. Limit buy and sell orders are arranged by their prices. A market order arrives and consumes limit orders of the same quantity of shares (the lightest region).

inventory to meet demand, as the electricity generated from renewable sources is subject to significant uncontrollable fluctuations (e.g., wind speed, weather conditions). However, as noted by Henriot (2014), Schmalensee (2012) and references therein, the trading volume in intraday markets is relatively low, and such illiquidity leads to high price impacts after trading. Thus, the power producer faces a dynamic optimization problem that minimizes the sum of three costs: the trading cost, the production cost, and the imbalance cost caused by the difference between demand and supply (defined as the electricity produced plus the net trading position).

### 1.2. Main Contribution and Literature Review

The main contribution of this paper is threefold: (1) We provide a partitioning algorithm for a class of linear quadratic MDPs with linear inequality constraints and nonconvex stagewise cost; see Section 2. (2) To the best of our knowledge, we are the first to introduce stochastic market depth as a Markov chain into the LOB optimal execution problem. Consequently, our model can incorporate different types of order activities, namely, the submission of limit orders outside the best quotes and order cancellation; see Section 3.2. In numerical examples, the stochastic policy significantly outperforms the deterministic strategy, especially for “unexpected” changes in the market; see Section 3.6. (3) For the renewable electricity management problem, the partitioning algorithm allows us to derive the exact optimal policy, which substantially outperforms the existing approximate solution (based on the relaxation of constraints).

Next we review the literature related to the methodology of this paper. The comparison to studies of both applications is deferred to their respective sections.

Our formulation of the MDP is closely related to the Markovian-jump linear quadratic (MLQ) optimal control proposed in Blair and Sworder (1975) and later extended by Chizeck et al. (1986). However, in the existing literature, the system is either stochastic but unconstrained as in MLQ or deterministic with constraints without the need for Bellman's principle of optimality (e.g., Bemporad et al. 2002). Other papers, such as Moore et al. (1999), consider more complicated noise structure; typically, however, only approximate/suboptimal solutions can be derived. This paper contributes to this stream of literature by presenting a partitioning algorithm to explicitly solve linear-quadratic problems that are stochastic and constrained.

Second, the stagewise cost in our paper can be nonconvex, while the realized total cost must be pathwise convex (see Assumption 1 and Equation (4)). Our setting relaxes the convexity assumption in multistage quadratic stochastic programs (Louveaux 1980, Birge and Louveaux 2011, Moore et al. 1999), although their formulation is more general in other dimensions, including non-Markovian dynamics. The relaxation of convexity is necessary for both applications in this paper. The handling of nonconvex cost subject to an even weaker convexity condition in quadratic programming is highlighted in Chen et al. (2014), in which an investor attempts to unwind a portfolio of multiple assets during a single period with a constant trading rate. By comparison, in our application to the optimal order execution, we consider liquidation of a single asset in multiple periods and explicitly model the LOB. This paper is organized as follows: A class of MDPs and the partitioning algorithm are introduced in Section 2. We present two applications in Sections 3 and 4. Section 5 provides concluding remarks and suggestions for future research.

## 2. A Class of Markov Decision Processes

In this section, we propose a partitioning algorithm for a class of linear-quadratic Markov decision processes. The formulation and various extensions are presented in Section 2.1. In Sections 2.2 and 2.3, we show the partitioning algorithm that solves the MDP: The state space can be partitioned into polyhedral regions; in each region, the optimal policy and the value function are closed-form linear and quadratic functions, respectively, of the state. In Section 2.4, the complexity of the algorithm is analyzed.

### 2.1. Formulation

Over a discrete-time finite horizon  $\{0, 1, \dots, n\}$ , consider an underlying Markov chain  $\{S_i\}_{i=0}^n$  governed by a transition probability matrix  $P^{(i)}$ , where  $S_i \in \{s_1, \dots, s_l\}$ . The MDP has a state vector  $\mathbf{x}_i \in \mathbb{R}^m$  and an action (decision) vector  $\mathbf{u}_i \in \mathbb{R}^p$ . To differentiate,

$S_i$  is referred to as the *exogenous* state, and  $\mathbf{x}_i$  as the *endogenous* state. The objective is to solve the following:

$$\begin{aligned} \min_{\{\mathbf{u}_i\}_{i=0}^n} & \mathbb{E} \left[ \sum_{i=0}^n \begin{pmatrix} \mathbf{u}_i \\ \mathbf{x}_i \end{pmatrix}' C_{S_i, i} \begin{pmatrix} \mathbf{u}_i \\ \mathbf{x}_i \end{pmatrix} \right] \\ \text{s.t.} & F'_{S_i, i} \mathbf{x}_i \leq D'_{S_i, i} \mathbf{u}_i \leq G'_{S_i, i} \mathbf{x}_i, \quad i = 0, \dots, n, \\ & \mathbf{x}_{i+1} = A_{S_i, i} \mathbf{x}_i + B_{S_i, i} \mathbf{u}_i, \quad i = 0, \dots, n-1. \end{aligned} \quad (1)$$

Here, the policy is assumed to be Markovian and nonanticipating, i.e.,  $\mathbf{u}_i$  is a function of  $\mathbf{x}_i$  and  $S_i$ . All coefficient matrices  $A \in \mathbb{R}^{m \times m}$ ,  $B \in \mathbb{R}^{m \times p}$ ,  $C \in \mathbb{R}^{(m+p) \times (m+p)}$ ,  $D \in \mathbb{R}^p$  and  $F, G \in \mathbb{R}^m$  can depend on stage  $i$  and the realization of the exogenous state  $S_i$ . The Markov chain can represent the status of the system (e.g., the market condition, the economic cycle). The dependence of the coefficient matrices on time and the exogenous state is arbitrary, and not necessarily linear or quadratic. Next, we show that (1) can be transformed into the following problem, in which the action is a scalar:

**Lemma 1.** *Problem (1) is equivalent to*

$$\begin{aligned} \min_{\{\hat{\mathbf{u}}_i\}_{i=0}^{\hat{n}}} & \mathbb{E} \left[ \sum_{i=0}^{\hat{n}} \begin{pmatrix} \hat{\mathbf{u}}_i \\ \hat{\mathbf{x}}_i \end{pmatrix}' \hat{C}_{\hat{S}_i, i} \begin{pmatrix} \hat{\mathbf{u}}_i \\ \hat{\mathbf{x}}_i \end{pmatrix} \right] \\ \text{s.t.} & \hat{F}'_{\hat{S}_i, i} \hat{\mathbf{x}}_i \leq \hat{\mathbf{u}}_i \leq \hat{G}'_{\hat{S}_i, i} \hat{\mathbf{x}}_i, \quad i = 0, \dots, \hat{n}, \\ & \hat{\mathbf{x}}_{i+1} = \hat{A}_{\hat{S}_i, i} \hat{\mathbf{x}}_i + \hat{B}_{\hat{S}_i, i} \hat{\mathbf{u}}_i, \quad i = 0, \dots, \hat{n}-1, \end{aligned} \quad (2)$$

where  $\hat{A} \in \mathbb{R}^{\hat{m} \times \hat{m}}$ ,  $\hat{B} \in \mathbb{R}^{\hat{m}}$ ,  $\hat{C} \in \mathbb{R}^{(\hat{m}+1) \times (\hat{m}+1)}$ , and  $\hat{F}, \hat{G} \in \mathbb{R}^{\hat{m}}$ .

The proof is presented in Online Appendix A. The coefficient matrices and dimensions in (2) are not necessarily the same as in (1) after the transformation; in particular,  $\hat{m} = m + p - 1$  and  $\hat{n} = np$ . The intuition of Lemma 1 is that a multivariate action can be activated in succession by dividing a stage into consecutive sub-stages and by allowing for one component of the action vector at each substage. As a result, we consider only the problem with a *scalar* action variable henceforth.

Using the framework for dynamic programming, we let  $J_i(\mathbf{x}_i, S_i)$  and  $u_i^*(\mathbf{x}_i, S_i)$  be the value function and the optimal policy, and we define the value of a state-action pair as

$$\begin{aligned} V_i(\mathbf{x}_i, S_i, u_i) &= \begin{pmatrix} u_i \\ \mathbf{x}_i \end{pmatrix}' C_{S_i, i} \begin{pmatrix} u_i \\ \mathbf{x}_i \end{pmatrix} \\ &+ \mathbb{E}_i [J_{i+1}(A_{S_i, i} \mathbf{x}_i + B_{S_i, i} u_i, S_{i+1}) | S_i]. \end{aligned} \quad (3)$$

By convention, we set  $V_i = J_i = +\infty$  for the infeasible endogenous state and  $J_{n+1} = 0$ . The optimality condition is given by the Bellman equation  $J_i(\mathbf{x}_i, S_i) = \min_{F'_{S_i, i} \mathbf{x}_i \leq u_i \leq G'_{S_i, i} \mathbf{x}_i} \{V_i(\mathbf{x}_i, S_i, u_i)\}$ . The following assumption is imposed throughout the paper:

**Assumption 1.** *Value function  $V_i$  is strictly convex in  $u_i$  for any  $(\mathbf{x}_i, S_i) \in \mathbb{R}^m \times \{s_1, \dots, s_l\}$  and  $i = 0, \dots, n$ .*

Several comments are in order for formulation (2). First, Assumption 1 is weaker than the usual assumptions that the coefficient matrix of the stagewise cost  $C$

must be positive definite for each  $i$  as in, e.g., Chizeck et al. (1986) and Birge and Louveaux (2011). As shown in Sections 3.4 and 4, for the two applications of interest here,  $C$  is never positive definite, while Assumption 1 holds under mild conditions. A sufficient condition for Assumption 1 is the so-called *pathwise convexity condition*: For any realized sample path of the Markov chain  $\{S_i = s_{k_i}\}_{i=0}^n$ , the realized quadratic cost

$$\sum_{i=0}^n \left( \begin{matrix} u_i \\ \mathbf{x}_i(u_0, \dots, u_{i-1}) \end{matrix} \right)' C_{s_{k_i}, i} \left( \begin{matrix} u_i \\ \mathbf{x}_i(u_0, \dots, u_{i-1}) \end{matrix} \right) \quad (4)$$

is jointly convex in  $(u_0, \dots, u_n)$ ,

after we represent  $\mathbf{x}_i$  as a linear function of  $(u_0, \dots, u_{i-1})$  and  $\mathbf{x}_0$ . Although the pathwise convexity condition is still stronger than Chen et al. (2014), in which the objective can be nonconvex pathwise, it is satisfied in our applications. Second, the quadratic cost and linear dynamics in (2) can be generalized to include constant terms and discrete additive random noise. More precisely, consider

$$\begin{aligned} \min_{\{u_i\}_{i=0}^n} & \mathbb{E} \left[ \sum_{i=0}^n \left( \begin{matrix} u_i \\ \mathbf{x}_i \end{matrix} \right)' C_{S_i, i} \left( \begin{matrix} u_i \\ \mathbf{x}_i \end{matrix} \right) + \mathbf{c}'_{S_i, i} \left( \begin{matrix} u_i \\ \mathbf{x}_i \end{matrix} \right) \right] \quad (5) \\ \text{s.t.} & F'_{S_i, i} \mathbf{x}_i + f_{S_i, i} \leq u_i \leq G'_{S_i, i} \mathbf{x}_i + g_{S_i, i}, \quad i = 0, \dots, n, \\ & \mathbf{x}_{i+1} = A_{S_i, i} \mathbf{x}_i + B_{S_i, i} u_i + d_{S_i, i}, \quad i = 0, \dots, n-1, \end{aligned}$$

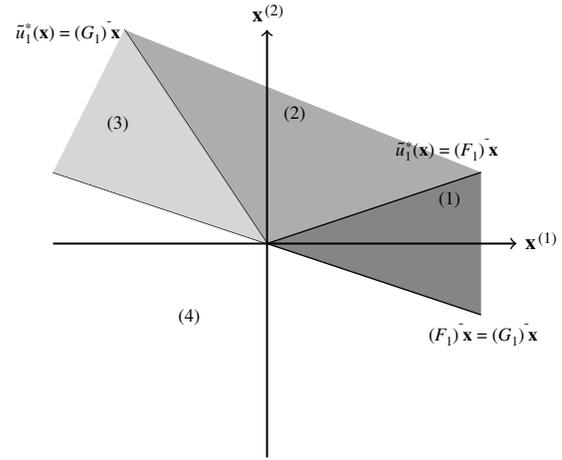
where  $f$ ,  $g$ , and  $d$  are scalars and  $\mathbf{c}$  is a vector of length  $m+1$ . Here,  $d$  represents the additive Markovian noise.<sup>1</sup> This formulation can be easily reduced<sup>2</sup> to (2) by appending a constant endogenous state variable, i.e.,  $(\mathbf{x}, 1)$ . For example,  $F'_{S_i, i} \mathbf{x}_i + f_{S_i, i}$  can be rewritten as  $(F_{S_i, i}, f_{S_i, i})'(\mathbf{x}_i, 1)$ .<sup>3</sup>

## 2.2. A Simple Case: The One-Step Deterministic Problem

To understand the intuition of the algorithm and shed light on the structure of the solution to the general case, we present the analytical solution of Problem (2) for the simplest case  $n = l = 1$ , i.e., the policy consists of two actions  $u_0$  and  $u_1$  while the Markov chain is a deterministic path. Because  $l = 1$ , we omit the subscript for the exogenous state.

We first analyze the last stage  $i = 1$ . The problem is infeasible for any endogenous state in the interior of the region  $G'_1 \mathbf{x}_1 \leq F'_1 \mathbf{x}_1$ . For any feasible state, the cost of the last stage can be expressed as  $a_1 u_1^2 + b'_1 \mathbf{x}_1 u_1 + \mathbf{x}'_1 c_1 \mathbf{x}_1 \triangleq (u_1, \mathbf{x}_1) C_1(u_1, \mathbf{x}_1)'$ , where  $a_1 \in \mathbb{R}$ ,  $b_1 \in \mathbb{R}^m$  and  $c_1 \in \mathbb{R}^{m \times m}$ . Because  $a_1 > 0$  by Assumption 1, the unconstrained minimizer is  $\tilde{u}_1^*(\mathbf{x}_1) = -b'_1 \mathbf{x}_1 / 2a_1$ , a linear function of  $\mathbf{x}_1$ . We partition the feasible endogenous state space by the linear boundaries  $\tilde{u}_1^* \leq F'_1 \mathbf{x}_1$ ,  $F'_1 \mathbf{x}_1 \leq \tilde{u}_1^* \leq G'_1 \mathbf{x}_1$  and  $G'_1 \mathbf{x}_1 \leq \tilde{u}_1^*$ . In each region, we can find the optimal control policy  $u_1^* = F'_1 \mathbf{x}_1$ ,  $u_1^* = \tilde{u}_1^*$ , and  $u_1^* = G'_1 \mathbf{x}_1$ , respectively, and obtain the quadratic value function accordingly. Therefore, the state space at  $i = 1$  is partitioned into four

**Figure 2.** Partition for the Last Stage of the Simple Case When  $m = 2$



*Notes.* In region (1),  $\tilde{u}_1^*(\mathbf{x}) \leq (F_1)' \mathbf{x} \leq (G_1)' \mathbf{x}$ ; hence  $u_1^* = (F_1)' \mathbf{x}$ . In region (2),  $(F_1)' \mathbf{x} \leq \tilde{u}_1^*(\mathbf{x}) \leq (G_1)' \mathbf{x}$ ; hence  $u_1^* = \tilde{u}_1^*(\mathbf{x})$ . In region (3),  $(F_1)' \mathbf{x} \leq (G_1)' \mathbf{x} \leq \tilde{u}_1^*(\mathbf{x})$ ; hence  $u_1^* = (G_1)' \mathbf{x}$ . In region (4), it is infeasible because  $(F_1)' \mathbf{x} > (G_1)' \mathbf{x}$ .

regions (including the infeasible one), as illustrated in Figure 2.

Surprisingly, a notable feature of the value function is its *differentiability* across boundaries<sup>4</sup> except those between feasible and infeasible regions, which can be verified by simple algebra. This differentiability is crucial: It guarantees that the value of a state-action pair  $V_0$  as defined in (3) does not have a sharp “wedge” in  $u_i$ , and thus, the minimum can be attained only at a stationary point or either constraint  $F'_0 \mathbf{x}$  or  $G'_0 \mathbf{x}$ .

Now, consider a state  $\mathbf{x}_0$  at  $i = 0$  and its optimal action  $u_0^*(\mathbf{x}_0)$ . The state-action pair  $(\mathbf{x}_0, u_0^*)$  will cause  $\mathbf{x}_1 = A_0 \mathbf{x}_0 + B_0 u_0^*$  to fall into one of the regions at  $i = 1$ . For instance, consider the region  $\mathcal{R}_1 \triangleq \{-b'_1 \mathbf{x}_1 / (2a_1) \leq F'_1 \mathbf{x}_1 \leq G'_1 \mathbf{x}_1\}$  in which  $\mathbf{x}_1$  falls. Then  $(\mathbf{x}_0, u_0^*)$  must satisfy  $(2a_1 F_1 + b_1)'(A_0 \mathbf{x}_0 + B_0 u_0^*) \geq 0$  and  $(G_1 - F_1)'(A_0 \mathbf{x}_0 + B_0 u_0^*) \geq 0$  so that  $\mathbf{x}_1 \in \mathcal{R}_1$ . Note that the value function of  $\mathcal{R}_1$  is  $J_1(\mathbf{x}_1) = \mathbf{x}'_1 (a_1 F_1 F_1' + b_1 F_1' + c_1) \mathbf{x}_1$ . Therefore, to verify the Bellman equation,  $(\mathbf{x}_0, u_0^*)$  must solve  $\min_{F'_0 \mathbf{x}_0 \leq u_0 \leq G'_0 \mathbf{x}_0} \{a_0 u_0^2 + b'_0 \mathbf{x}_0 u_0 + \mathbf{x}'_0 c_0 \mathbf{x}_0\}$ , where  $a_0 u_0^2 + b'_0 \mathbf{x}_0 u_0 + \mathbf{x}'_0 c_0 \mathbf{x}_0 \triangleq (u_0, \mathbf{x}_0) C_0(u_0, \mathbf{x}_0)'$  and  $J_1(A_0 \mathbf{x}_0 + B_0 u_0)$ . We denote the unconstrained minimizer of  $a_0 u_0^2 + b'_0 \mathbf{x}_0 u_0 + \mathbf{x}'_0 c_0 \mathbf{x}_0$  as  $\tilde{u}_0^*$ , i.e.,  $\tilde{u}_0^* = -b'_0 \mathbf{x}_0 / (2a_0)$ . Similar to the case of  $i = 1$ ,  $u_0^*$  depends on the relative positions of  $\tilde{u}_0^*$ ,  $F'_0 \mathbf{x}_0$  and  $G'_0 \mathbf{x}_0$ , leading to the following linear boundaries:  $\tilde{u}_0^* \leq F'_0 \mathbf{x}_0 \leq G'_0 \mathbf{x}_0$ ,  $F'_0 \mathbf{x}_0 \leq \tilde{u}_0^* \leq G'_0 \mathbf{x}_0$ , and  $F'_0 \mathbf{x}_0 \leq G'_0 \mathbf{x}_0 \leq \tilde{u}_0^*$ . Combining this with the condition that  $\mathbf{x}_1 \in \mathcal{R}_1$ , we obtain three possible regions in which  $\mathbf{x}_0$  may fall, as well as the corresponding  $u_0^*$  and value function:

(1)  $(2a_1 F_1 + b_1)'(A_0 + B_0 F'_0) \mathbf{x}_0 \geq 0$ ,  $(G_1 - F_1)'(A_0 + B_0 F'_0) \mathbf{x}_0 \geq 0$ ,  $-b'_0 \mathbf{x}_0 / (2a_0) \leq F'_0 \mathbf{x}_0 \leq G'_0 \mathbf{x}_0$ ;  $u_0^*(\mathbf{x}_0) = F'_0 \mathbf{x}_0$ ,  $J_0(\mathbf{x}_0) = \mathbf{x}'_0 (a_0 F_0 F_0' + b_0 F_0' + c_0) \mathbf{x}_0$ .

(2)  $(2a_1 F_1 + b_1)'(A_0 - B_0 b'_0 / (2a_0)) \mathbf{x}_0 \geq 0$ ,  $(G_1 - F_1)'(A_0 - B_0 b'_0 / (2a_0)) \mathbf{x}_0 \geq 0$ ,  $F'_0 \mathbf{x}_0 \leq -(b'_0 \mathbf{x}_0 / 2a_0) \leq G'_0 \mathbf{x}_0$ ;  $u_0^*(\mathbf{x}_0) = (b_0 / 2a_0)' \mathbf{x}_0$ ,  $J_0(\mathbf{x}_0) = \mathbf{x}'_0 (-b_0 b'_0 / 4a_0 + c_0) \mathbf{x}_0$ .

(3)  $(2a_1F_1 + b_1)'(A_0 + B_0G'_0)\mathbf{x}_0 \geq 0$ ,  $(G_1 - F_1)'(A_0 + B_0G'_0)\mathbf{x}_0 \geq 0$ ,  $-b'_0\mathbf{x}_0/(2a_0) \geq G'_0\mathbf{x}_0 \geq F'_0\mathbf{x}_0$ ;  $u_0^*(\mathbf{x}_0) = G'_0\mathbf{x}_0$ ,  $J_0(\mathbf{x}_0) = \mathbf{x}'_0(a_0G_0G'_0 + b_0G'_0 + c_0)\mathbf{x}_0$ .

In each region, the first two linear boundaries correspond to the condition that  $\mathbf{x}_1 \in \mathcal{R}_1$ ; the third linear boundary corresponds to the positions of  $\tilde{u}_0^*$ ,  $F'_0\mathbf{x}_0$  and  $G'_0\mathbf{x}_0$ . Note that the region may be a null set if the half spaces do not intersect.

Given that we have obtained three feasible regions for the state  $\mathbf{x}_0$  that lead  $\mathbf{x}_1$  to fall in  $\mathcal{R}_1$ , the same process is then conducted for other regions,  $\mathcal{R}_2 \triangleq \{F'_1\mathbf{x}_1 \leq -b'_1\mathbf{x}_1/(2a_1) \leq G'_1\mathbf{x}_1\}$  and  $\mathcal{R}_3 \triangleq \{G'_1\mathbf{x}_1 \leq -b'_1\mathbf{x}_1/(2a_1), F'_1\mathbf{x}_1 \leq G'_1\mathbf{x}_1\}$  at  $i = 1$ .

The solution highlights the key steps in generating linear boundaries and partitioning the state space by backward induction. The region at  $i + 1$  is mapped linearly into the state space at  $i$  due to the linear state dynamics. The region at  $i$  is then further partitioned by comparison of the unconstrained minimizer and constraints, both linear functions of the endogenous state as the Bellman equation yields a linear unconstrained minimizer.

### 2.3. The General Case

The special case of  $n = l = 1$  (recall that  $n$  is the length of the horizon and  $l$  is the number of exogenous states) gives insights into the proof for the general case. However, several technical difficulties must be addressed. For example, why do the regions cover the whole state space and not intersect with one another? How is the differentiability inherited through backward induction? Why are the quadratic coefficients ( $a_0$  and  $a_1$  in Section 2.2) always positive? Theorem 1 gives a formal statement; the proof is available in Online Appendix A. Algorithm 1 can be used to recursively find the partitions as well as the optimal policy and value function.

**Theorem 1.** For stage  $i$  and exogenous state  $S_i = s_j$ , we can find a partition of  $\mathbb{R}^m$  consisting of  $n_{j,i}$  polyhedral regions (of linear boundaries),  $\{\mathcal{P}_r^{j,i}\}_{r=1}^{n_{j,i}}$ , such that for  $\mathbf{x}_i \in \mathcal{P}_r^{j,i}$ , the optimal policy and value function of (2) are of the form

$$u_i^*(\mathbf{x}_i, s_j) = (M_r^{j,i})'\mathbf{x}_i, \quad J_i(\mathbf{x}_i, s_j) = \mathbf{x}'_i N_r^{j,i} \mathbf{x}_i,$$

or  $J_i(\mathbf{x}_i, s_j) = +\infty$  if  $\mathbf{x}_i$  is infeasible, i.e., either  $(F_{s_j,i} - G_{s_j,i})'\mathbf{x}_i > 0$  or  $\mathbf{x}_{i+1}$  has a positive probability of being infeasible at  $i + 1$  for all  $u_i$ . The coefficients  $M_r^{j,i}$  of the optimal policy and the coefficients  $N_r^{j,i}$  of the value function can be recursively computed. Moreover,  $J_i(\cdot, s_j)$  is differentiable at  $\mathbf{x}_i$  if  $J_i$  is feasible in a neighborhood of  $\mathbf{x}_i$ .

Two comments are in order. First, the value function and optimal policy in different regions can be recursively computed. We briefly describe the intuition. In the Bellman equation, for stage  $i$  and exogenous

state  $s_j$ , the unconstrained minimizer  $\tilde{u}_i^*$  is given by a univariate quadratic program:

$$\begin{aligned} \tilde{u}_i^* &= \arg \min_{u_i \in \mathbb{R}} \left\{ \begin{pmatrix} u_i \\ \mathbf{x}_i \end{pmatrix}' C_{s_j,i} \begin{pmatrix} u_i \\ \mathbf{x}_i \end{pmatrix} + \sum_{k=1}^l P_{jk}^{(i)} \mathbf{x}'_{i+1} N_{r_k}^{k,i+1} \mathbf{x}_{i+1} \right\} \\ &= \arg \min_{u_i \in \mathbb{R}} \{a_{s_j,i} u_i^2 + b'_{s_j,i} \mathbf{x}_i u_i + \mathbf{x}'_i c_{s_j,i} \mathbf{x}_i\} \\ &= -b'_{s_j,i} \mathbf{x}_i / (2a_{s_j,i}), \end{aligned}$$

provided the structure of the partition at  $i + 1$  and that  $\mathbf{x}_{i+1}$  falls into  $\mathcal{P}_k^{k,i+1}$  for  $k = 1, \dots, l$ . Then the constrained minimizer  $u_i^*$  can be the following:  $F'_{s_j,i} \mathbf{x}_i$  if  $\tilde{u}_i^* \leq F'_{s_j,i} \mathbf{x}_i$ ,  $\tilde{u}_i^*$  if  $F'_{s_j,i} \mathbf{x}_i \leq \tilde{u}_i^* \leq G'_{s_j,i} \mathbf{x}_i$ ,  $G'_{s_j,i} \mathbf{x}_i$  if  $G'_{s_j,i} \mathbf{x}_i \leq \tilde{u}_i^*$ , or infeasible. In all cases,  $u_i^*$  (if feasible) is a linear function of  $\mathbf{x}_i$ . Combining this with the linear inequalities that  $A_{s_j,i} \mathbf{x}_i + B_{s_j,i} u_i^*$  must fall into the intersection of  $\mathcal{P}_k^{k,i+1}$  for  $k = 1, \dots, l$ , we obtain polyhedral regions in the state space at  $i$ . Finally, the coefficient of the linear policy  $u_i^*$  is  $M_r^{j,i} = F_{s_j,i,r} - b_{s_j,i,r} / (2a_{s_j,i})$  or  $G_{s_j,i,r}$  depending on whether the constraints are active, and the coefficient of the value function is  $N_r^{j,i} = a_{s_j,i} (M_r^{j,i})' M_r^{j,i} + b_{s_j,i} M_r^{j,i} + c_{s_j,i}$ . Hence, backward induction can proceed.

Second, the differentiability of the value function is crucial in recursively computing the partitions, which guarantees that the maximum is only attained at stationary points when solving the Bellman equation. This property can essentially be attributed to the differentiability of  $h(\mathbf{x}) \triangleq \min_{u \leq f' \mathbf{x}} \{a u^2 + b' \mathbf{x} u + \mathbf{x}' c \mathbf{x}\}$  for  $a > 0$ , i.e., the switching of  $h(\mathbf{x})$  on the boundary  $f' \mathbf{x} = -b' \mathbf{x} / (2a)$  is smooth. Such a comparison of the unconstrained minimizer and the constraint is the only source of new linear boundaries at stage  $i$ , as other linear boundaries are inherited from stage  $i + 1$ . Therefore, partitioning will not cause nondifferentiability in backward induction.

### 2.4. Complexity

Next we investigate the number of regions in a partition of the general formulation (2) as a function of problem dimensions  $m$ ,  $n$  and  $l$ , which we refer to as the complexity. Note that the partitioning algorithm is an offline algorithm. The partitions are to be computed once, and the optimal policy for any realization of the Markov chain will follow. This is particularly important for optimal execution problems (Section 3), in which the order execution must be re-solved every day.

Although one can easily obtain a theoretical upper bound  $3^{(n+1)l^{n+1}}$  for the number of regions,<sup>5</sup> it greatly overestimates, as similarly noted by Bemporad et al. (2002). Many linear boundaries derived from the Bellman equation, in the form of half-spaces, do not intersect to form a region and are thus redundant. For example, in the LOB application, the number of regions is bounded above by  $(2l)^{n+1}$ , in which  $m = 2$  and  $l, n \in \mathbb{Z}_+$ . For electricity management in intraday markets (Section 4), the complexity is bounded above by  $2l^n$ , in which  $m = 3$ ,  $l = 4$  and  $n \in \mathbb{Z}_+$ .<sup>6</sup>

**Algorithm 1** (The Partitioning Algorithm)

- 1: **initialization**
- 2: set  $i \leftarrow n$
- 3: **for**  $j \leftarrow 1$  to  $l$  **do**
- 4: compute the cost function of the last stage  $a_{s_j,n}u_n^2 + (b_{s_j,n})'x_n u_n + x_n'c_{s_j,n}x_n \triangleq (u_n, x_n)C_{s_j,n}(u_n, x_n)'$ , and the unconstrained minimizer  $\tilde{u}_n^* = -(b_{s_j,n})'x_n / (2a_{s_j,n})$
- 5: partition the state space into four regions:  $(F_{s_j,n} - G_{s_j,n})'x_n \geq 0$  (the infeasible region),  $-(b_{s_j,n})'x_n / (2a_{s_j,n}) \leq (F_{s_j,n})'x_n \leq (G_{s_j,n})'x_n$ ,  $(F_{s_j,n})'x_n \leq -(b_{s_j,n})'x_n / (2a_{s_j,n}) \leq (G_{s_j,n})'x_n$ , and  $(F_{s_j,n})'x_n \leq (G_{s_j,n})'x_n \leq -(b_{s_j,n})'x_n / (2a_{s_j,n})$
- 6: compute the optimal policy  $u_n^*(x_n)$  in the above regions: infeasible,  $(F_{s_j,n})'x_n$ ,  $-(b_{s_j,n})'x_n / (2a_{s_j,n})$ , and  $(G_{s_j,n})'x_n$
- 7: compute the value function in the above regions by letting  $u_n = u_n^*$  in  $a_{s_j,n}u_n^2 + (b_{s_j,n})'x_n u_n + x_n'c_{s_j,n}x_n$
- 8: **end for**
- 9: **for**  $i \leftarrow n - 1$  to 0 **do**
- 10: suppose the linear boundaries of all regions in state  $s_j$  at stage  $i + 1$  are  $\{l_{i+1,j}^{(k)}\}_{k=1}^{r_{i+1,j}}$ ; combine all linear boundaries  $\cup_{j=1}^l \{l_{i+1,j}^{(k)}\}_{k=1}^{r_{i+1,j}}$  and eliminate redundant boundaries to obtain a refined partition
- 11: **for**  $j \leftarrow 1$  to  $l$  **do**
- 12: **for all region**  $\mathcal{P}$  in the refined partition **do**
- 13: **if**  $\mathcal{P}$  is infeasible for some  $S_{i+1} = s_k$  and  $P_{jk}^{(i)} > 0$  **then**
- 14: go to Step 12
- 15: **end if**
- 16: compute  $a_{s_j,i}u_i^2 + (b_{s_j,i})'x_i u_i + x_i'c_{s_j,i}x_i \triangleq V_i(x_i, s_j, u_i)$  according to (3) by using the quadratic value function in region  $\mathcal{P}$  for all  $S_{i+1} = s_k$  in the Bellman equation
- 17: define three regions according to the unconstrained minimizer  $\tilde{u}_i^* = -(b_{s_j,i})'x_i / (2a_{s_j,i})$ :  
(1)  $A_{s_j,i}x_i + B_{s_j,i}(F_{s_j,i})'x_i \in \mathcal{P}$ ,  $\tilde{u}_i^* \leq (F_{s_j,i})'x_i$ ; (2)  $A_{s_j,i}x_i + B_{s_j,i}\tilde{u}_i^* \in \mathcal{P}$ ,  $(F_{s_j,i})'x_i \leq \tilde{u}_i^* \leq (G_{s_j,i})'x_i$ ; and (3)  $A_{s_j,i}x_i + B_{s_j,i}(G_{s_j,i})'x_i \in \mathcal{P}$ ,  $(G_{s_j,i})'x_i \leq \tilde{u}_i^*$ , where the optimal policy is  $u_i^* = (F_{s_j,i})'x_i$ ,  $u_i^* = \tilde{u}_i^*$ , and  $u_i^* = (G_{s_j,i})'x_i$ , respectively, in each region
- 18: compute the value function by letting  $u_i = u_i^*$  in  $a_{s_j,i}u_i^2 + (b_{s_j,i})'x_i u_i + x_i'c_{s_j,i}x_i$
- 19: **end for**
- 20: flag the rest state space as infeasible
- 21: **end for**
- 22: **end for**

**2.4.1. Randomized Computational Experiments.** Because the theoretical bound is overestimating, we conduct randomized computational experiments to investigate the complexity as a function of problem dimensions  $m$ ,  $n$ , and  $l$ . For each set of dimensions, we solve 1,000 instances of the MDPs by randomly generating their coefficient matrices and the transition probability matrix of the Markov chain (see Online Appendix B for details). Then the average number of regions at  $i = 0$  over all instances and the standard error are computed. The results are shown in the left panel of Table 1 for different combinations of dimensions. By comparison, we also randomly generate 1,000 sets of parameters for both applications and evaluate the average complexity.<sup>7</sup> The results are shown in the right panel of Table 1.

Two comments are in order on the complexity of the algorithm and Table 1. First, when  $m = 2$ , the

partitioning algorithm can handle large  $n$  and  $l$ . This is because linear boundaries on  $\mathbb{R}^2$  generate a linear number of regions. In the high-frequency example in Table 4 (in Section 3.6), there are only 2,592 regions for  $m = 2$ ,  $n = 500$ , and  $l = 51$ . When  $m = 3$ , for  $l \leq 3$ , the complexity increases mildly in  $n$ ; however, for  $l = 4$ , the number of regions grows fast as  $n$  increases. In Table 1, the dimension  $(m, n, l) = (3, 3, 4)$  yields an average complexity of 1,270.2, much larger than the rest. This is partly because of more extreme instances in higher dimensions: more than 10 of them have a complexity of more than 10,000, giving rise to the high standard error. Second, the complexity of the applications (right panel) is significantly lower than the randomized computational experiments (left panel) for the same dimension. We conjecture that two factors might play a role in practice. First, the coefficient matrices are usually sparse. In the randomized experiments, components

**Table 1.** Average Number of Regions That Are Feasible in a Partition and the Standard Error (in Parentheses) for Combinations of  $(l, m, n)$  in the Randomized Computational Experiments (Left Panel) and the Two Applications (Right Panel) for 1,000 Instances Each

$m$	$l$	Randomized experiments				Real applications			
		$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 0$	$n = 1$	$n = 2$	$n = 3$
2	1	3.0 (0.0)	5.1 (0.0)	7.1 (0.0)	9.2 (0.0)	3.0 (0.0)	4.0 (0.0)	5.0 (0.0)	6.0 (0.0)
2	2	3.0 (0.0)	6.0 (0.0)	11.7 (0.1)	21.4 (0.2)	3.0 (0.0)	4.7 (0.0)	6.5 (0.0)	8.3 (0.0)
2	3	3.0 (0.0)	6.8 (0.1)	16.3 (0.2)	36.5 (0.5)	3.0 (0.0)	5.5 (0.0)	8.1 (0.0)	10.8 (0.1)
3	1	3.0 (0.0)	5.1 (0.0)	7.0 (0.0)	8.9 (0.0)	LOB application above			
3	2	3.0 (0.0)	5.7 (0.0)	10.4 (0.1)	18.7 (0.3)				
3	3	3.0 (0.0)	6.5 (0.1)	14.3 (0.2)	56.3 (6.1)	Electricity application below			
3	4	3.0 (0.0)	7.1 (0.1)	18.6 (0.3)	1,270.2 (744.1)				

Notes. The application to LOBs corresponds to the top three rows in the right panel. The application to electricity management corresponds to the row of  $m = 3$  and  $l = 4$ .

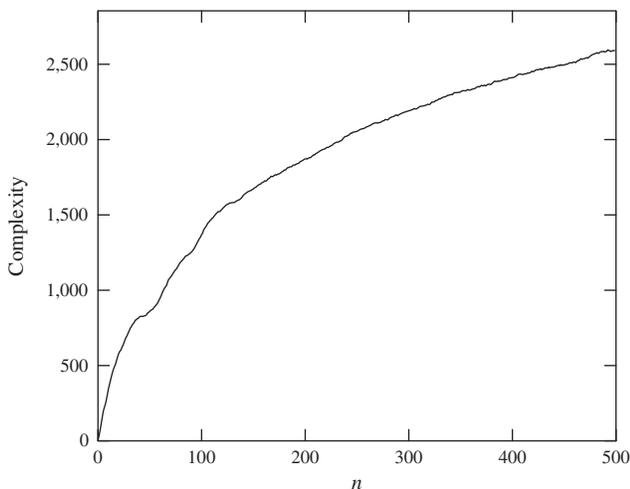
of  $B$ ,  $C$ , and  $G$  are nonzero, whereas in applications, only a few states and actions have interactions. For example, in the application to optimal order execution,  $C$  has only three nonzero entries. Sparsity might lead to simplified dynamics and lower complexity. Second, there are often implicit structures embedded in real problems, and regions that do not satisfy the structural constraints are eliminated. For example, in the LOB application, the optimal policy is increasing in the exogeneous state, i.e., the market depth. Such monotonicity might lead to sublinear growth in complexity in  $n$ , as illustrated in Figure 3.

**2.4.2. Curse of Dimensionality.** The curse of dimensionality is a serious issue for dynamic optimization. When the dimension of the state space increases, the memory requirement and the computation time usually scale exponentially. In formulation (2), if one resorts to grid-based numerical schemes (e.g., discretizing the

state space and performing value iteration), then the required memory and computation time increase exponentially in  $m$ . Thus, such numerical schemes suffer from the curse of dimensionality. By contrast, the complexity of the partitioning algorithm can be quite insensitive to the dimension  $m$  even for large  $m$ , partly because there is no loop involving  $m$  in Algorithm 1. In particular, the following example shows that the algorithm works well for  $m = 10$ , in which  $l$  is small.

**Example 1** (Inventory Management with a Stochastic Lead Time). Consider a finite-horizon, discrete-time inventory model with a backlogged demand and a stochastic lead time. The model is a variant of that in Section 9.6 of Zipkin (2000), in which we have quadratic costs instead of their piecewise linear costs. More precisely, the lead time is either 1 or 10 with equal probability, independent of everything else. The deterministic demand is 1 in each period. At the beginning of each period, the inventory manager makes an order knowing the realization of its stochastic lead time (but the realization of the lead time of future orders is unknown). In period  $i$ , let  $u_i \geq 0$  be the order size (a decision variable);  $x_i$  be the net inventory position before ordering;  $\{d_{i,1}, \dots, d_{i,9}\}$  be the “pipeline” vector, i.e., the order of size  $d_{i,j}$  will arrive right before period  $i + j$ . The holding and backlog costs are assumed to be quadratic and symmetric, i.e., the cost from  $i$  to  $i + 1$  is  $(x_i - 1)^2$ , as demand is always 1. Moreover, there is a quadratic order cost  $u_i^2$ . To formulate the inventory management problem in the form of (2), let  $S_i \in \{1, 10\}$  be the exogenous state, where  $\{S_i\}_{i=0}^n$  are independent Bernoulli random variables, and the dynamics of the endogenous state be  $x_{i+1} = x_i - 1 + d_{i,1} + 1_{S_i=1}u_i$ ,  $d_{i+1,j} = d_{i,j+1}$  for  $j = 1, \dots, 8$ , and  $d_{i+1,9} = 1_{S_i=10}u_i$ . We consider the optimal decision of making three orders ( $n = 3$ ), and the total cost is  $\sum_{i=0}^2 (u_i^2 + (x_i - 1)^2) + \sum_{i=3}^{12} (x_i - 1)^2$ , where the second sum accounts for the holding and backlog costs in the next 10 periods before the pipeline order is

**Figure 3.** Complexity as a Function of  $n$  for the LOB Application, When  $m = 2$  and  $l = 51$



Note. The parameters have the same values as Figure 6.

**Table 2.** Comparison of Computation Time Between the Partitioning Algorithm and the Grid-Based Scheme

Example	Dimension ( $m$ )	Partitioning	Grid-based	
			Error level	
			$\approx 0.1\%$	$\approx 1\%$
Low-frequency trading (Table 4)	2	0.6 s	12 m 11 s	33 s
High-frequency trading (Table 4)	2	6 h	173 h	4 h 40 m
Electricity management ( $n = 3$ in Table 6)	3	3 m 50 s	5 h 36 m	3 m 29 s
Electricity management ( $n = 4$ in Table 6)	3	9 h 46 m	36 h 19 m	42 m 88 s
Inventory (Example 1, Section 2.4.2)	10	17.8 s	3 h 18 m (Err = 84.7%)	
Multiple LOB (Example 2, Section 3.7)	4	1 h 41 m	44 h (Err = 5.6%)	

Notes. The error is defined to be  $|\hat{J}_0(\mathbf{x}_0, s_0) - J_0(\mathbf{x}_0, s_0)|/J_0(\mathbf{x}_0, s_0)$ , where  $\hat{J}$  is the value function computed from the grid-based scheme and  $(\mathbf{x}_0, s_0)$  is the initial state in the examples. For examples of low dimensions ( $m = 2$  and  $m = 3$ ), we report the computation time of the grid-based scheme for two approximate error levels (0.1%, finer grid) and (1%, coarser grid). For examples of high dimensions ( $m = 4$  and  $m = 10$ ), the grid-based scheme is tested on a grid as fine as the memory can handle ( $5^{10}$  and  $100 \times 40 \times 40 \times 40$ , respectively). For all numerical examples in this paper, calculations are performed on a laptop with a 2.40 GHz Intel i7-4700 CPU and 8 GB memory.

cleared, i.e., when  $d_{i,j} = 0$  for all  $j$ . The matrix form of this formulation is available in Online Appendix B.

As mentioned in Section 9.6.3.3 of Zipkin (2000), for inventory models with a stochastic lead time, the structure of the optimal policy is usually unknown. Although some heuristics and approximate algorithms are proposed (see, e.g., Ehrhardt 1984, Anupindi et al. 1996, Bradley and Robinson 2005), none of them provides an analytical solution. Moreover, a large lead time leads to a longer pipeline, or equivalently, higher dimension of the state space, which makes numerical solution infeasible. For Example 1, the dimension is  $m = 10$  (not counting the constant state) and a grid-based scheme would first discretize each state variable, e.g., into 10 grid points with uniform steps. Such discretization leads to a 10-dimensional grid of  $10^{10}$  points. Note that 10 grid points for each state variable would still be too coarse in most cases. Yet value iteration is virtually infeasible for such a large-scale grid: Regardless of the memory requirement ( $>100$  GB), the computation would be very difficult even for a single iteration. By contrast, the closed-form solution given by the partitioning algorithm has a simple structure: The partition on date 0 has 74 regions, and it can be computed in less than 20 seconds on an average CPU.

To conclude Section 2.4, we list, in Table 2, the computation time of the examples appearing in this paper. For all examples, the partitioning algorithm outperforms the grid-based scheme in terms of computation time if the latter aims to achieve 0.1% error (see the notes of Table 2 for definition). In particular, examples 1 and 2 have state spaces of high dimensions ( $m = 10$  and  $m = 4$ ). For the computation of grid-based schemes to be feasible, the grid cannot be too fine, and hence, the discretization error is large. Their analytical solutions have only 74 and 2,319 regions, respectively, given by the partitioning algorithm.

### 3. Application One: Limit Order Book Optimal Execution

Currently, most equity and derivative exchanges around the world either are electronic limit order markets only or have electronic limit orders in addition to on-exchange market making (Parlour and Seppi 2008). The prevalence of LOBs has motivated an extensive body of research; see, e.g., Parlour and Seppi (2008), O'Hara (1995), and Gould et al. (2013) for comprehensive surveys. In particular, for theoretical studies, see, e.g., Kyle (1985), Glosten and Milgrom (1985), Roşu (2009), Goettler et al. (2005), Foucault (1999), Foucault et al. (2005), Cont et al. (2010), and Guo et al. (2013); for empirical studies, see, e.g., Biais et al. (1995), Ahn et al. (2001), Rinaldo (2004), and Cao et al. (2008).

We propose a limit order book model with stochastic market depth, to which the partitioning algorithm can be applied; the optimal order execution can then be solved in a closed form. In the literature on optimal order execution, the LOB dynamics may or may not be explicitly specified. For example, in Bertsimas and Lo (1998), Almgren and Chriss (2001), Almgren (2012), Chen et al. (2014), Guo and Zervos (2015), the problem is solved by assuming some price impact functions *without* directly modeling the LOB. By comparison, Obizhaeva and Wang (2013), Alfonsi et al. (2010) solve the optimal execution problem for a one-sided LOB; Tsoukalas et al. (2016) study the problem in two-sided and block-shaped LOBs of multiple correlated assets; Horst and Naujokat (2014) consider a more general objective, tracking a target function, for a two-sided, block-shaped LOB. Predoiu et al. (2011) study a one-sided LOB with a general shape and resilience function whose market depth remains static, and thus extend the results of Obizhaeva and Wang (2013), Alfonsi et al. (2010), Fruth et al. (2014) who propose an LOB model with time-varying but deterministic market depth. We assume that the LOB is block shaped

and that the market depth is stochastic, following a discrete-time Markov chain.

As a feature of our model, the stochastic nature of market depth is supported by empirical evidence. For example, based on 33 Hang Seng Index component stocks, Ahn et al. (2001) show that the market depth is stochastic, with the average exhibiting a reverse U-shape in a day, i.e., the market depth peaks around noon; more precisely, the maximal market depth in a day (at noon) can be more than 20% higher than the average for the whole day; the minimal market depth (in the morning) also deviates –40% from the average. Moreover, static/deterministic market depth is inconsistent with some order activities. For example, placement of limit orders at and outside the best quotes (i.e., non-quote-improving limit orders) and order cancellation lead to stochastic changes in market depth. Biais et al. (1995) empirically show that these two types of order activities may account for more than 60% of limit order activities. In our model, market depth can be reverse U-shaped in expectation with some stochastic variation; non-quote-improving limit order placement and cancellation can also be incorporated. As pointed out in, e.g., Kempf and Korn (1999), Rinaldo (2004), Cao et al. (2008), Kavajecz (1999), and Goettler et al. (2009), market depth is an important indicator for LOBs. Therefore, our model provides a more realistic treatment of optimal order execution.

The stochastic model that we use to describe market depth is a discrete-time Markov chain. Markovian and semi-Markovian models are often used to describe LOB dynamics. For example, Guilbaud and Pham (2013) propose an LOB model whose spread follows a Markov chain in which the optimal market-making problem is solved by dynamic programming; in Fodra and Pham (2015a, b), the price return and tick time are modeled by a Markov renewal process, and various stylized facts, such as the mean reversion of the price return, can be captured. Modeling the market depth as a Markov chain allows us to formulate the optimal order execution problem as an MDP. A comparison of LOB models is provided in Table 3.

This section is organized as follows. After introducing the model in Section 3.1, we show in Section 3.2 that the model can incorporate different order activities. The optimal execution problem is formulated as an MDP in Section 3.3. In Section 3.4 we demonstrate that the optimal execution problem fits into the framework in Section 2. We show various empirical implications of the model in Section 3.5. We conduct numerical

experiments in Section 3.6 and illustrate the advantage of our model over models with deterministic market depth. The model is extended to include multiple trading venues in Section 3.7.

### 3.1. The Limit Order Book Dynamics

We consider the ask side of the LOB. At time  $i$ , the one-sided order book is fully determined by the following variables:

- $M_i$ : The fundamental value of the stock;
- $d_i$ : The spread between  $M_i$  and the best ask price,  $d_i \geq 0$ ;
- $Q_i$ : The market depth of the book.

Similar to Obizhaeva and Wang (2013), Fruth et al. (2014), and Tsoukalas et al. (2016), the LOB in our model is of a block shape: The block is on  $[M_i + d_i, \infty)$ , with a height of  $Q_i$ . More precisely, all limit orders are uniformly placed on the price interval  $[M_i + d_i, \infty)$ , and the market depth  $Q_i$  indicates the average number of limit order shares per unit price.

We take snapshots of tick-by-tick buy transactions in the LOB during the period  $[0, T]$ ; empirical studies of LOBs are usually based on tick-by-tick data (e.g., Biais et al. 1995). More precisely, transactions are executed at  $0 = t_0 < t_1 < \dots < t_n = T$ , with the order size  $u_i$  at  $t_i$ . The transactions consume the limit orders on the book from low prices to high prices. In response, the spread is pushed up at  $t_i$ :  $M_{i+} = M_i$ ,  $d_{i+} = d_i + u_i/q_i$ ,  $Q_{i+} = Q_i$ . Because we are interested only in discrete time, we use  $\mathcal{F}_i$  to represent all the information available at  $t_i$ .

Next, we specify the dynamics of  $M_i$ ,  $d_i$  and  $Q_i$  for  $i = 0, 1, \dots, n$ . Following the standard literature, we assume  $M_t$  to be a martingale:  $\mathbb{E}[M_{i+1} | \mathcal{F}_i] = M_i$ . As we shall see in Section 3.3, because of the martingale property, the optimal trading policy does not involve  $M_t$ .

To describe the dynamics of the spread between  $t_i$  and  $t_{i+1}$ , we assume that the price impact decays over time and exponentially mean-reverts to 0, governed by a time-dependent resilience rate  $\rho_i > 0$  (see, e.g., Fruth et al. 2014, Alfonsi et al. 2010). More precisely, the evolution of  $d_i$  follows  $d_{i+1} = e^{-\rho_i(t_{i+1}-t_i)} d_{i+} = e^{-\rho_i(t_{i+1}-t_i)} (d_i + u_i/Q_i)$ . The mean reversion of the spread is well documented in empirical studies (see, e.g., Biais et al. 1995).

A key feature of our model is that the market depth  $Q_i$  is stochastic. We assume that  $Q_i$  follows a finite state Markov chain, as Markovian or semi-Markovian LOB models can capture various stylized facts and provide analytical tractability (Fodra and Pham 2015a, b; Guilbaud and Pham 2013). The Markov chain is constructed on  $(\Omega, \mathcal{F}_i, \mathbb{P})$ , independent of  $M_i$ , with exogenous states  $0 < s_1 < s_2 < \dots < s_l$ . The transition probability matrix  $P^{(i)} \in \mathbb{R}^{l \times l}$  with  $\mathbb{P}(Q_{i+1} = s_k | Q_i = s_j) = P_{jk}^{(i)}$  can be time-inhomogeneous. Our subsequent analytical results *do not require* particular functional forms of  $P^{(i)}$ . In numerical examples, however, we focus on

**Table 3.** Summary of Papers Related to LOB Modeling

Predoiu et al. (2011)	General shape, static market depth
Fruth et al. (2014)	Block shape, dynamic but deterministic market depth
This paper	Block shape, stochastic market depth

Geometric Ornstein-Uhlenbeck processes, as in Almgren (2012), and take the transition probability to be

$$P_{jk}^{(i)} \propto \exp\left(-\frac{(s_k - s_j - \theta(\mu_i - s_j)(t_{i+1} - t_i))^2}{2\sigma^2 s_j^2 (t_{i+1} - t_i)}\right), \quad (6)$$

where  $\mu_i \in \mathbb{R}$ ,  $\sigma, \theta \in \mathbb{R}^+$  are the transient mean, mean-reversion rate, and volatility of the process, respectively.

In summary, given the LOB state  $(M_i, d_i, Q_i)$  and transaction sizes  $\{u_i\}$ , the following events are triggered. The fundamental value  $M_i$  and the market depth  $Q_i$  do not instantaneously change, while the spread has a jump at  $t_i$ . During  $(t_i, t_{i+1}]$ ,  $M_i$  has a mean-zero increment. The spread mean-reverts to zero as determined by an exponential rate  $\rho_i$ . The market depth follows a discrete Markov chain.

### 3.2. Different Order Activities

Our model can incorporate order activities mentioned in Biais et al. (1995). Snapshots of the LOBs (relative to  $M_t$ ) at  $t_i$  and  $t_{i+1}$  are illustrated in Figure 4. The following order activities are considered:

*Activity 1.* Market buy orders submitted at time  $t_i$ ;

*Activity 2.* Quote-improving limit sell orders placed between  $t_i$  and  $t_{i+1}$ ;

*Activity 3.* Non-quote-improving limit sell orders placed between  $t_i$  and  $t_{i+1}$ ;

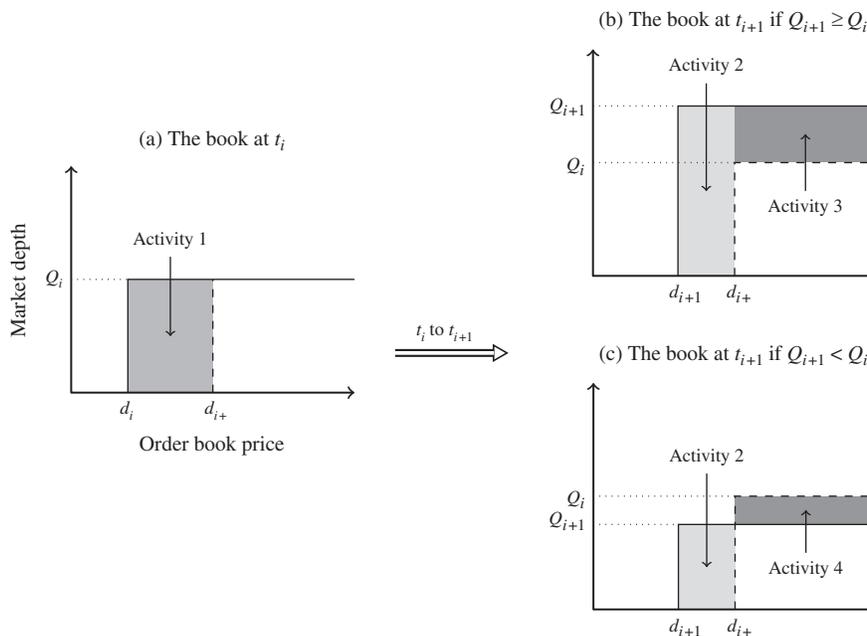
*Activity 4.* Cancellation of limit orders between  $t_i$  and  $t_{i+1}$ .

At time  $t_i$ , the best ask price is driven up by a market buy transaction, as shown in panel (a) of Figure 4. This is caused by Activity 1: A market order of size  $u_i$  will consume the same amount of shares on the book and drives up the best ask price from  $M_i + d_i$  to  $M_i + d_{i+} = M_i + d_i + u_i/Q_i$ .

Note that the spread always improves during  $(t_i, t_{i+1})$  because of mean reversion. Hence, the book can change from panel (a) to panel (b) or (c) as in Figure 4, depending on whether  $Q_{i+1} \geq Q_i$  or  $Q_{i+1} < Q_i$ . In both cases, Activity 2 means that limit orders are submitted below  $d_{i+}$  during  $(t_i, t_{i+1})$ ; more precisely, the size of quote-improving limit orders during the period is indicated by the area of the lightly shadowed block:  $Q_{i+1}(d_{i+} - d_{i+1})$ .

If  $Q_{i+1} \geq Q_i$  (panel (b)), the darkly shadowed block can only be caused by the placement of non-quote-improving limit sell orders, which are limit orders placed above the spread  $d_{i+}$  during  $(t_i, t_{i+1})$ , i.e., Activity 3. The net amount of non-quote-improving limit orders placed per unit price during  $(t_i, t_{i+1})$  is  $Q_{i+1} - Q_i$ . If  $Q_{i+1} < Q_i$  (panel (c)), the decrease in the market depth can only be caused by limit order cancellation (Activity 4). The net cancellation of  $Q_i - Q_{i+1}$  non-quote-improving limit orders per unit price during  $(t_i, t_{i+1})$  leads to the darkly shadowed block in panel (c). Meanwhile, Activity 2 is represented by the lightly shadowed block, as in panel (b).

Figure 4. Different Order Activities in Our LOB Model



Notes. At  $t_i$ , a transaction (Activity 1) happens and drives up the best ask price as a result, as shown in panel (a); the dashed line is the spread right after the trade. During  $(t_i, t_{i+1})$ , the order book resilience causes the spread to revert from  $d_{i+}$  to  $d_{i+1}$ , outlined by the dashed lines and solid lines, respectively, in the right panels. Meanwhile,  $Q_{i+1}$  can be either larger than  $Q_i$  (panel (b)) or less than  $Q_i$  (panel (c)). The shadowed blocks are caused by corresponding order activities 2, 3, and 4.

For optimal execution models with static or deterministic market depth, only Activities 1 and 2 are incorporated. Thanks to stochastic market depth, Activities 3 and 4, namely, the submission of non-quote-improving limit orders and order cancellation, can also be incorporated into our model. As shown in Biais et al. (1995), Activities 3 and 4 may account for more than 60% of the total limit order activities.

### 3.3. Optimal Order Execution

In this section, we consider the optimal execution problem (see, e.g., Predoiu et al. 2011, Fruth et al. 2014, Alfonsi et al. 2010): A risk-neutral institutional trader wants to buy  $x_0$  shares of stock from the LOB specified in Section 3.1 by submitting market orders, i.e., Activity 1 in Section 3.2 is conducted by the trader, while Activities 2–4 are performed by other market participants. Because the average price per share for the transaction at  $t_i$  is  $M_i + d_i + u_i/(2Q_i)$ , the trader's objective is to minimize the expected overall execution cost  $\mathbb{E}[\sum_{i=0}^n u_i(M_i + d_i + u_i/(2Q_i))]$ , where  $\sum_{i=0}^n u_i = x_0$ . Note that it is possible to include a penalty term for risk in the objective, such as the mean variance formulation (e.g., Tse et al. 2013) or the mean quadratic variation formulation (Forsyth et al. 2012). However, in optimal order execution problems for LOBs, it is standard to assume the trader to be risk-neutral and minimize the expected cost without risk adjustment (e.g., Bertsimas and Lo 1998, Predoiu et al. 2011). The main reason for this is that optimal executions are only part of the overall asset management program, and an institutional trader may want to hold a stock for a longer period  $T'$  (e.g., several months) than the execution period  $T$ . The volatility from period  $T'$  tends to be much larger than that from  $T$ .

Because we focus only on the ask side of the book, we impose the constraints  $u_i \geq 0$ . Moreover, the trader is allowed to make decisions by relying on all information available. Therefore, the admissible policy set of the trader is given by  $\Theta = \{u_0, u_1, \dots, u_n \mid \sum_{i=0}^n u_i = x_0, u_i \geq 0, u_i \text{ is } \mathcal{F}_i\text{-measurable}\}$ . The optimal order execution problem can thus be formulated as

$$\min_{\{u_i\}_{i=0}^n \in \Theta} \mathbb{E} \left[ \sum_{i=0}^n u_i \left( M_i + d_i + \frac{u_i}{2Q_i} \right) \right]. \quad (7)$$

Let  $x_i = x_0 - \sum_{k=0}^{i-1} u_k$  be the state variable representing the remaining shares to fulfill right before the trade at  $t_i$ . Because of the structure of the objective function and the dynamics, the problem is a typical MDP. For simplicity, we assume equally spaced trading time, i.e.,  $t_i = iT/n = i\Delta t$  in the following. Let  $\bar{u}_i^*(x_i, d_i, Q_i, M_i)$  be the optimal policy of (7) and  $\bar{J}_i(x_i, d_i, Q_i, M_i) := \mathbb{E}[\sum_{k=i}^n \bar{u}_k^*(M_k + d_k + \bar{u}_k^*/(2Q_k)) \mid x_i, d_i, Q_i, M_i]$  be the value function at stage  $i$ .

As shown in, e.g., Bertsimas and Lo (1998) and Predoiu et al. (2011), the fundamental value of stock  $M$  does not appear in their optimal execution strategies, at least for additive price impact models. The claim also holds in our model. More precisely, setting  $M_t \equiv 0$ , we have the following:

$$\min_{\{u_i\}_{i=0}^n \in \Theta} \mathbb{E} \left[ \sum_{i=0}^n u_i \left( d_i + \frac{u_i}{2Q_i} \right) \right]. \quad (8)$$

Similarly, let  $u_i^*(x_i, d_i, Q_i)$  and  $J_i(x_i, d_i, Q_i)$  be the optimal policy and the value function of (8) at stage  $i$ .

**Proposition 1.** (i) *The optimal policy of (7) is equivalent to that of (8):  $\bar{u}_i^*(x_i, d_i, Q_i, M_i) = u_i^*(x_i, d_i, Q_i)$ .*

(ii) *The optimal execution costs differ by a constant:  $\bar{J}_0(x_0, d_0, Q_0, M_0) = J_0(x_0, d_0, Q_0) + x_0 M_0$ .*

(iii) *The optimal execution cost decreases in resilience:  $J'_0(x_0, d_0, Q_0) \leq J_0(x_0, d_0, Q_0)$ , if the resilience rate  $\rho'_i \geq \rho_i$  for all  $i$ , ceteris paribus.*

(iv) *Assume that the transition probability  $P_{jk}^{(i)}$  has the following property: For  $s_{j_1} > s_{j_2}$ , there exists  $a(s_{j_1}, s_{j_2}) > 0$  such that*

$$\begin{cases} \mathbb{P}(Q_{i+1} = s_k \mid Q_i = s_{j_1}) \geq \mathbb{P}(Q_{i+1} = s_k \mid Q_i = s_{j_2}) & \text{if } s_k \geq a, \\ \mathbb{P}(Q_{i+1} = s_k \mid Q_i = s_{j_1}) < \mathbb{P}(Q_{i+1} = s_k \mid Q_i = s_{j_2}) & \text{if } s_k < a. \end{cases}$$

*Then, the optimal execution cost decreases in market depth:  $J_0(x_0, d_0, Q'_0) \leq J_0(x_0, d_0, Q_0)$ , if  $Q'_0 \geq Q_0$ .*

### 3.4. The Analytical Solution

We can transform (8) into the general form (2) by simply letting  $S_i = Q_i$ ,  $\mathbf{x}_i = (x_i, d_i)'$ ,

$$\begin{aligned} A_{s_j, i} &= \begin{pmatrix} 1 & 0 \\ 0 & e^{-\rho_i \Delta t} \end{pmatrix}, & B_{s_j, i} &= \begin{pmatrix} -1, & \frac{e^{-\rho_i \Delta t}}{s_j} \end{pmatrix}' \\ C_{s_j, i} &= \begin{pmatrix} 1/(2s_j) & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{pmatrix}, \end{aligned}$$

and  $F_{s_j, i} = \mathbf{0}$ . The equality constraint  $\sum_{i=0}^n u_i = x_0$  can be translated into an inequality constraint  $u_i \leq x_i$  or, equivalently,  $G_{s_j, i} = (1, 0)'$  at each stage. This is a sufficient and necessary condition for the problem to be feasible at the next stage. Moreover, at the final stage  $t_n$ , the optimal policy is simply assigned to  $u_n^* = x_n$ , as all remaining shares will be fulfilled.

Note that the cost coefficient matrix  $C$  is never positive definite: The eigenvalues of  $C_{s_j, i}$  are  $(\sqrt{4s_j^2 + 1} + 1)/4s_j$ ,  $0$ , and  $(1 - \sqrt{4s_j^2 + 1})/4s_j$ . Therefore,  $u_i(d_i + u_i/2Q_i)$  is not convex in  $(u_i, x_i, d_i)$  for any  $Q_i$ . Nevertheless, the following proposition guarantees the convexity of  $V_i$  in  $u_i$  and, thus, Assumption 1.

**Proposition 2.** *If  $Q_i > \exp(-2\rho_i \Delta t)Q_{i+1}$  almost surely for all  $i = 0, 1, \dots, n-1$ , then  $V_i(x_i, d_i, Q_i, u_i)$  defined in (3) is strictly convex in  $u_i$ .*

The condition will be satisfied if the variation in market depth is not too large ( $Q_{i+1}/Q_i$  is small), the resilience effect is strong ( $\rho$  is large) or the trader does not trade very frequently ( $\Delta t$  is large). If the condition is not satisfied, then price manipulation as defined in Huberman and Stanzl (2004) may exist: One can buy at low market depth to push the spread high and then sell at high market depth (as if there is no nonnegativity constraint) to make infinite profit. Intuitively, the condition rules out price manipulation and guarantees that a global minimum exists, i.e.,  $V_i$  is convex in  $u_i$ . Therefore, Assumption 1 enhances the set of practical problems that the algorithm can solve. Now we can apply Algorithm 1 and solve the optimal execution problem for LOBs with stochastic market depth.

The structure of the optimal execution strategy is implied by Theorem 1: At stage  $i$  for some realized  $Q_i$ , the  $(x, d)$  space (we consider only the  $\mathbb{R}^+ \times \mathbb{R}^+$  quadrant) is partitioned by a set of linear boundaries  $\{x = a_k d \mid 0 < a_1 < a_2 < \dots < a_K\}$ . For small  $a_k$ , the price is relatively high compared to the remaining shares. In these regions, the optimal policy is to execute 0 shares at the current stage and wait for the price reversion. However, there might be more than one region in which the optimal policy  $u_i$  is 0. For example, for  $a_1 d_i \leq x_i \leq a_2 d_i$ , it may be optimal for the trader to wait at  $t_i$  and trade at  $t_{i+1}$ , while for  $x_i \leq a_1 d_i$ , the trader should wait longer. Different future decisions are reflected in the partition of the current state as well.

### 3.5. Empirical Implications

Our model and the resulting optimal execution policy can provide an explanation for some stylized facts suggested in the empirical literature.

(1) *Stochastic market depth.* Ahn et al. (2001) find that the average intraday market depth has a reverse U-shaped pattern. To capture this feature, we can choose particular forms of  $\mu_i$  and  $\sigma$  so that on average  $Q_i$  first increases and then decreases. One choice is  $\mu_i = a Q_0 \sin(t_i \pi / T)$ , where  $a > 0$  and  $T = 1$  day, leading to reverse U-shaped market depth that mimics the empirical observations in Ahn et al. (2001).

(2) *Market depth and order execution.* Rinaldo (2004), Cao et al. (2008) find that when the market depth is high, traders are more willing to execute market orders. Our model can provide an explanation for this phenomenon by showing that the price impact and total execution cost are decreasing in market depth. Because large trading cost and price impact are unfavorable, traders tend to execute more orders when the market depth is higher. Indeed, the price impact is given by

$$d_i - d_0 = \left( \exp\left(-\sum_{k=0}^{i-1} \rho_k \Delta t\right) - 1 \right) d_0 + \sum_{j=0}^{i-1} \exp\left(-\sum_{k=j}^{i-1} \rho_k \Delta t\right) \frac{u_j}{Q_j}.$$

Therefore, the total execution cost of the trades  $(u_0, u_1, \dots, u_n)$  is

$$\begin{aligned} & \sum_{i=0}^n u_i \left( M_i + d_i + \frac{u_i}{Q_i} \right) \\ &= \sum_{i=0}^n M_i u_i + d_0 \sum_{i=0}^n \exp\left(-\sum_{k=0}^{i-1} \rho_k \Delta t\right) u_i \\ & \quad + \sum_{i=0}^n \frac{u_i^2}{2Q_i} + \sum_{j<i} \exp\left(-\sum_{k=j}^{i-1} \rho_k \Delta t\right) \frac{u_i u_j}{Q_j}. \end{aligned}$$

Both the price impact and total execution cost decrease in market depth  $Q_i$ . A similar result is shown in part (iv) of Proposition 1 in which the optimal execution cost also decreases in  $Q_0$ .

(3) *Nonlinear price impact.* Kempf and Korn (1999) observe that empirically the price impact is not a linear function of order size. Our model provides an explanation for this phenomenon: Although the price impact at time  $t_i$  is linear with coefficient  $1/Q_i$ , the coefficient is stochastic over time. Therefore, from a statistical perspective, fitting a linear model with constant coefficients will produce poor results. This phenomenon is inconsistent with block-shaped LOBs with static market depth.

(4) *Wait/buy region.* From an optimization perspective, a region in a partition at  $t_i$  corresponds to a set of active constraints  $u_{i_1} = u_{i_2} = \dots = u_{i_k} = 0$  for  $\{i_1, \dots, i_k\} \subset \{i, i+1, \dots, n\}$ , which depends on the realization of market depth. That is, for a realization of  $\{Q_j\}_{j=i}^n$ , the order executions starting from two endogenous states in the same region at  $t_i$  share the same pattern: They both buy ( $u_k > 0$ ) or both wait ( $u_k = 0$ ) for  $k \geq i$ . As an informal analogy, the partitioning process backward in time is a branching process of the decision tree. A region in the partition represents a “branch” of all future decisions extending to the final stage. In other words, although the system is stochastic, the partition informs whether the trader should buy or wait for any future realization of the Markov chain. Moreover, if two endogenous states differ in future decisions for some realization of the Markov chain, they belong to two regions in the partition of the current state space. This is a generalization of the concept of wait/buy regions in Fruth et al. (2014).

### 3.6. The Advantage of the Stochastic Policy

In the model presented in Section 3.3, the trader is allowed to exploit all available information when making a decision. However, if the trader uses only the information at  $t_0$ , i.e.,  $u_i \in \mathcal{F}_0$  are deterministic variables, then the following deterministic optimization problem arises:

$$\begin{aligned} & \min_{\{u_i\}_{i=0}^n} \mathbb{E} \left[ \sum_{i=0}^n u_i \left( d_i + \frac{u_i}{2Q_i} \right) \right] \\ & \text{s.t. } \sum_{i=0}^n u_i = x_0, \quad u_i \geq 0, \\ & \quad d_{i+1} = e^{-\rho_i \Delta t} \left( d_i + \frac{u_i}{Q_i} \right). \end{aligned} \tag{9}$$

To apply quadratic programming (QP), we compute  $\mathbb{E}[1/Q_i]$  given  $Q_0$ . (Because  $Q_i$  is a Markov chain and  $Q_0$  is known, one can compute  $\mathbb{P}(Q_i = s_j | Q_0)$  with the  $i$ -step transition matrix and, thus, the expectation.) Then, we express  $d_i$  in terms of  $u_0, \dots, u_{i-1}$  so that the objective is a quadratic function of  $\{u_i\}_{i=0}^n$  with known coefficients. This deterministic formulation is closely related to the model in Fruth et al. (2014). By Proposition 2, the quadratic objective function of (9) is positive definite if and only if  $\mathbb{E}[1/Q_{i+1}] > \exp(-2\rho_i\Delta t)\mathbb{E}[1/Q_i]$  for  $i = 0, 1, \dots, n - 1$ . Note that QP yields a suboptimal execution policy for the MDP formulation because the dynamics of the Markov chain  $\{Q_i\}_{i=0}^n$  are not included in the decision making. Thus, the gap between the deterministic and the MDP formulation is narrowed when the randomness diminishes. In summary, we have the following result:

**Proposition 3.** (i) *The optimal value of (9) is an upper bound for that of the MDP formulation (8) and can be solved by QP.*

(ii) *Suppose that the transition matrix of  $Q_i$  is given by (6). If the volatility of the market depth goes to zero, then the gap between the optimal value of the deterministic and the MDP formulation converges to zero.*

We illustrate and compare the performance of the stochastic and deterministic optimal policies with a numerical example. We investigate cases of low- and high-frequency trading for the parameters  $\rho = 6$ ,  $T = 1$  day,  $d_0 = 0.1$ ,  $Q_0 = 500$  and  $x_0 = 10,000$ , while  $n$  equals 10 and 500, corresponding to one trade every 2,340 seconds (low frequency) and 23.4 seconds (high frequency), respectively. There are initially 500 units of limit order shares per unit price. We let the state space of the Markov chain be  $\{Q_0 e^s \mid s = -0.2, -0.1, \dots, 0.7, 0.8\}$  for  $n = 10$  and  $\{Q_0 e^s \mid s = -0.2, -0.18,$

$\dots, 0.78, 0.8\}$  for  $n = 500$ . The market depth follows the geometric Ornstein–Uhlenbeck process in (6). We choose  $\theta = 5$  and  $\mu_i = Q_0(1 + \sin(i\pi/n))$  to reproduce the reverse U-shaped average market depth.<sup>8</sup>

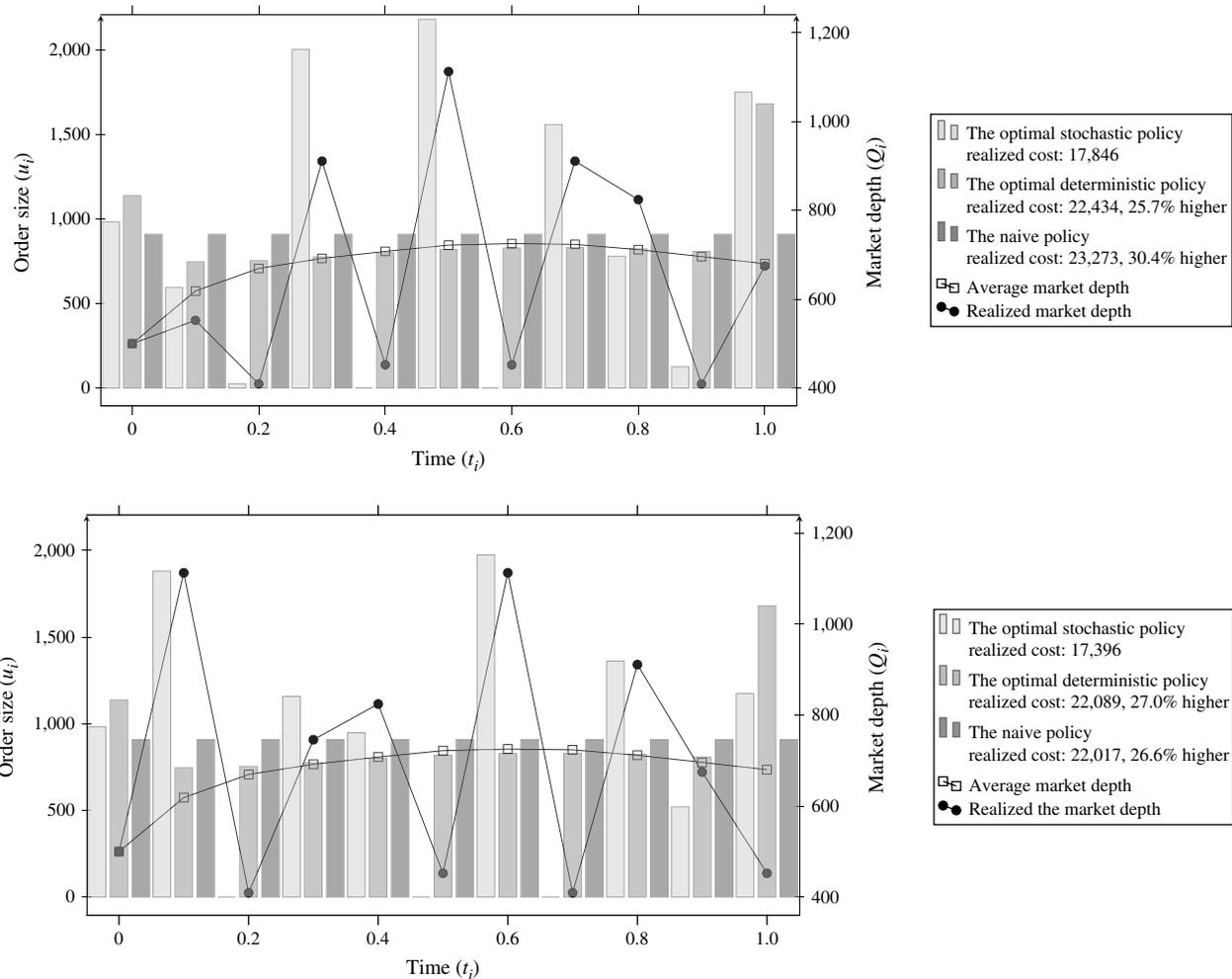
As a benchmark, we also test the performance of a naive policy that executes an even amount of  $x_0/(n + 1)$  shares per stage. We compute the theoretical expected execution costs of the three policies using the partitioning algorithm.<sup>9</sup> Moreover, we simulate 10,000 sample paths of  $Q_i$  according to (6). For each sample path, we use the three policies and obtain their realized execution costs. In Table 4, we compare summary statistics for the execution costs. We also illustrate “unexpected” sample paths of the market depth and the corresponding optimal execution policy in Figures 5 and 6. In summary, we have following observations:

(1) The average realized cost of the optimal deterministic policy is significantly higher than that of the stochastic policy. In the low-frequency case ( $n = 10$ ), the improvement is more than that of the optimal deterministic policy over the naive policy, which is very simple and involves no optimization. In the high-frequency case ( $n = 500$ ), the differences between the expected costs of the optimal stochastic and deterministic policies are 0.4%, which is less than the low-frequency case. This result is probably due to the smoothness of the Markov chain in high frequency (see Endnote 8). Nevertheless, as noted by Brogaard et al. (2014), high-frequency trading (HFT) is associated with daily revenues of approximately \$0.43 per \$10,000 traded. Therefore, the benefit of the stochastic policy over the deterministic policy in the high-frequency case, approximately \$40 per \$10,000 traded, remains substantial when measured in HFT terms. This result highlights the advantage of our LOB model: Without incorporating stochastic market depth, a trader who

**Table 4.** Expected Values, the Sample Averages of the Realized Costs over 10,000 Sample Paths, and Differences in the Percentage (from the Stochastic Policy) of the Three Policies

		Expected	Average	$p$ -value
$n = 10$	Stochastic	19,882	19,910	
$\sigma = 2$	Deterministic	(6.9%) 21,251	(7.0%) 21,273	$\leq 10^{-10}$
	Naive	(9.8%) 21,822	(9.9%) 21,846	
$n = 10$	Stochastic	19,598	19,627	
$\sigma = 3$	Deterministic	(8.2%) 21,197	(8.1%) 21,218	$\leq 10^{-10}$
	Naive	(11.1%) 21,764	(11.0%) 21,787	
$n = 500$	Stochastic	23,616	23,621	
$\sigma = 2$	Deterministic	(0.4%) 23,707	(0.4%) 23,712	$\leq 10^{-10}$
	Naive	(11.3%) 26,295	(11.3%) 26,311	
$n = 500$	Stochastic	23,920	23,926	
$\sigma = 3$	Deterministic	(0.4%) 24,017	(0.4%) 24,022	$\leq 10^{-10}$
	Naive	(11.3%) 26,631	(11.3%) 26,646	

*Notes.* We also conduct a paired-samples  $t$ -test for the costs of the stochastic and deterministic policies, with the null hypothesis that the two policies have costs with an equal mean. The  $p$ -value is significant, indicating that the stochastic policy is better than the deterministic policy.

**Figure 5.** Order Executions and Realized Costs of the Three Policies Under Two Unexpected Sample Paths in Low Frequency ( $n = 10, \sigma = 2$ )

mistakenly uses a deterministic policy, even the optimal one, will suffer substantial losses.

(2) When the market depth is more volatile (larger  $\sigma$ ), the optimal stochastic policy is more advantageous, especially in the low-frequency case.

(3) For the unexpected sample paths of the market depth, the stochastic policy outperforms the deterministic policy by a larger margin than that of the average cost. Adapting to the unexpected changes, the stochastic policy tends to trade in low volume when the market depth is low.

In practice, the realized market depth may deviate from the average, i.e., the reverse U-shape, from time to time. For instance, a corporate press release may greatly alter the limit order flow and trading activities and, thus, the market depth; the presence of other large traders (which may or may not be known to the public) can cause unusual market depth changes; and the stock market often observes different patterns on special dates, e.g., postholiday such as December 26.

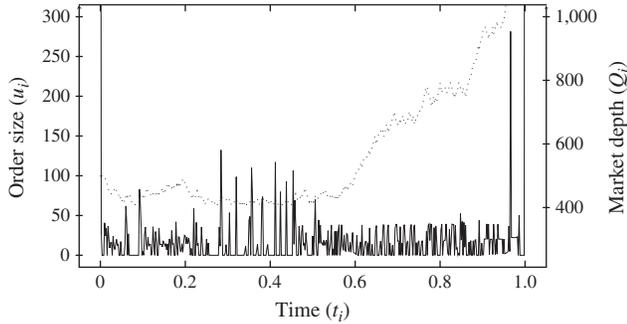
The stochastic policy has a significant advantage over the deterministic policy in those scenarios.

### 3.7. Extension to Multiple LOBs

A stock can usually be traded in multiple venues, including primary and regional exchanges; each venue maintains its own limit order book. To minimize the execution cost, an investor who wants to buy a large amount of shares must split the order not only over time but also among multiple venues. Our model can be extended to incorporate this case. Next, we briefly describe the formulation and discuss how to apply the partitioning algorithm for the closed-form optimal policy.

Consider the ask sides of  $m$  LOBs and trading dates  $t_i$  for  $i = 0, \dots, n$ . For simplicity, let  $t_{i+1} - t_i = \Delta t$ . At  $t_i$ , the LOBs are described by  $d_{i,j}$ , the best ask prices of the  $j$ th LOB, and  $Q_{i,j}$ , the market depth (assuming zero fundamental value). Here,  $S_i \triangleq (Q_{i,1}, \dots, Q_{i,m})$  is a Markov chain on  $\mathbb{R}^m$ . The investor has initial demand

**Figure 6.** Realized Order Executions of the Optimal (Stochastic) Policy Under a Particular Sample Path in the High-Frequency Case ( $n = 500, \sigma = 2$ )



Notes. The dotted curve represents the realized market depth. The solid line shows the realized stochastic policy. The order executions of the deterministic and naive policies (not shown in the figure) follow smooth and flat curves close to zero, in sharp contrast to the pattern of the stochastic policy. For this particular sample path, the deterministic and naive policies have execution costs 4.2% and 28.2% higher, respectively, than that of the stochastic policy.

$x_0$  and adaptively determines the size of the market order  $u_{i,j}$  that is submitted to the  $j$ th LOB at  $t_i$ .

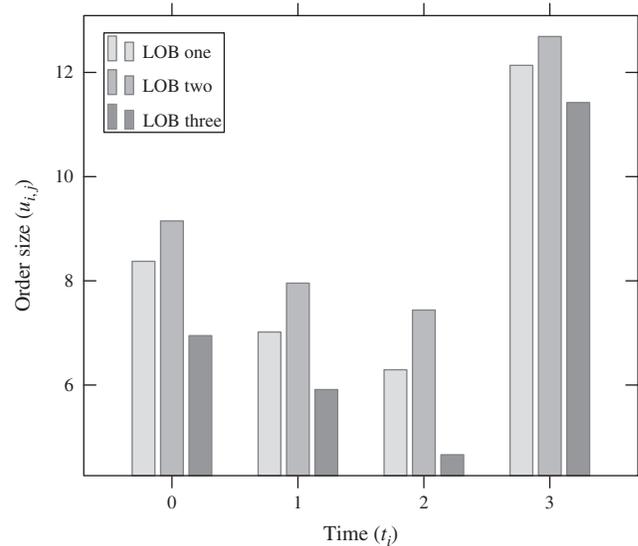
$$\begin{aligned} \min_{u_{i,j} \geq 0} \quad & \mathbb{E} \left[ \sum_{j=1}^m \sum_{i=0}^n u_{i,j} \left( d_{i,j} + \frac{u_{i,j}}{2Q_{i,j}} \right) \right] \\ \text{s.t.} \quad & x_{i+1} = x_i - \sum_{j=1}^m u_{i,j}, \quad d_{i+1,j} = e^{-\rho_{i,j}\Delta t} \left( d_{i,j} + \frac{u_{i,j}}{Q_{i,j}} \right), \\ & \sum_{j=1}^m u_{n,j} = x_n. \end{aligned}$$

According to Section 2.1, the  $m$ -dimensional decision variable  $\{u_{i,j}\}_{j=1}^m$  can be transformed to a scalar by making the investor decide  $u_{i,j}$  sequentially for  $j = 1, \dots, m$  at  $t_i$ . More precisely, the transformed problem has a length of horizon  $m(n+1)$ , a scalar decision variable  $\hat{u}_i$ , an endogenous state  $(\hat{x}_i, \hat{d}_{i,1}, \dots, \hat{d}_{i,m})$ , and an exogenous state  $\hat{S}_i \triangleq (\hat{Q}_{i,1}, \dots, \hat{Q}_{i,m})$ .

$$\begin{aligned} \min_{\hat{u}_i \geq 0} \quad & \mathbb{E} \left[ \sum_{i=1}^{m(n+1)} \hat{u}_i \left( \hat{d}_{i,j} + \frac{\hat{u}_i}{2\hat{Q}_{i,j}} \right) \right] \\ \text{s.t.} \quad & \hat{x}_{i+1} = \hat{x}_i - \hat{u}_i, \\ & \hat{d}_{i+1,j} = \begin{cases} e^{-\rho_{i,j}\Delta t} (\hat{d}_{i,j} + \hat{u}_i / \hat{Q}_{i,j}) & \text{if } j = i - km \text{ for some } k \in \mathbb{N}, \\ \hat{d}_{i,j} & \text{otherwise,} \end{cases} \\ & \hat{u}_{m(n+1)} = \hat{x}_{m(n+1)}, \\ & \hat{Q}_{i,j} = Q_{\lfloor (i-1)/m \rfloor, j} \end{aligned}$$

where  $\lfloor z \rfloor$  represents the largest integer less than equal to  $z$ . In the transformed problem, if  $i = km + j$ , then the investor focuses on the  $j$ th LOB, and one can interpret  $\hat{u}_i$  as  $u_{k,j}$ .  $\hat{S}_i$  is still a Markov chain: It “freezes” for  $m - 1$  periods and then transits every  $m$ th period with

**Figure 7.** Optimal Order Execution Policy for Three Limit Order Books Averaged over 1,000 Sample Paths of the Market Depth



the same probability as  $S_{\lfloor (i-1)/m \rfloor}$ . Then, the partitioning algorithm can be used to solve the optimal execution problem.<sup>10</sup>

When the market is fragmented (i.e., there are several trading venues), as is the case of the U.S. stock market, a grid-based scheme would suffer from the curse of dimensionality, and the computation time would increase in  $m$  at a tremendous rate. See Table 2 in Section 2.4.2 for comparison of computation time for the following example for the case of  $m = 4$ .

**Example 2.** Consider  $N = 3$  limit order books and  $t_i = i$  for  $i = 0, 1, 2, 3$  ( $\Delta t = 1$ ). The market depth of all three order books is equal ( $Q_{i,1} = Q_{i,2} = Q_{i,3} = Q_i$ ) and has two states  $Q_i \in \{10, 15\}$ . The transition probability matrix of the Markov chain  $Q_i$  has all components equal to 0.5. The different resilience rates are  $\rho_{i,1} = 0.71$ ,  $\rho_{i,2} = 0.86$  and  $\rho_{i,3} = 0.50$ . Initially, the investor has demand  $x_0 = 100$ ; the prices are  $d_{0,j} = 0$  for  $j = 1, 2, 3$ , and the initial market depth is  $Q_0 = 10$ .

For Example 2, the partitioning algorithm solves the optimal execution cost 74.13 in less than two hours. There are 2,319 regions at  $t = 0$ . The optimal order execution policy averaged over 1,000 sample paths of the market depth is illustrated in Figure 7.

#### 4. Application Two: Renewable Electricity Management in Intraday Markets

The renewable electricity management problem that we consider was first studied in Aïd et al. (2016): A power producer has some renewable energy sources

and thermal plants and manages energy in the intraday market in continuous time, with a linear price impact function similar to Almgren and Chriss (2001). At the end of the horizon, the producer determines whether to use the thermal plants to generate electricity to meet the unmatched demand. This introduces an inequality constraint, as production must be nonnegative. The authors then give an approximate solution based on the relaxation of the constraint. In this section, we present a discrete-time version of the model studied in Aid et al. (2016). With the partitioning algorithm, we can analytically and precisely solve the discrete-time version of the constrained problem. The optimal policy is then compared to the approximate solution proposed in Aid et al. (2016) by numerical examples in Section 4.2.

#### 4.1. The Discrete-Time Model

The agent has two sources of power production, i.e., renewable energy, which is uncontrollable and highly unpredictable, and thermal plants, whose production can be precisely controlled. Over the planning horizon, the agent generates power by renewable energy to meet the demand; we refer to the difference between the demand and the power generated as the residual demand. Meanwhile, the agent can trade in the intraday market to balance his or her position by future contracts expiring at terminal time  $T$ . At  $T$ , the agent decides whether to use thermal plants to produce power and then delivers the electricity from future contracts plus the thermal power to match the residual demand. The objective is to minimize the cost arising from trading, thermal power generation, and unmatched residual demand.

Aid et al. (2016) formulate the optimal trading/production problem in continuous time with two features that are worth mentioning. First, the price process and the residual demand are subject to random fluctuations, modeled by correlated Brownian motions. In discrete time, they can be approximated by binomial white noise. Second, the purchase/sale of the agent has a linear price impact on the intraday market. Price impact is a common factor in market microstructure to measure the impact of transactions on the market. The authors propose an approximate solution to the optimization problem.

Using the partitioning algorithm, we can solve the discrete-time equivalent in a closed form. The discrete-time problem is formulated as follows: Consider  $t_i = iT/n \triangleq i\Delta t$  for  $i = 0, \dots, n$ . The agent's decision variables are denoted by  $\{u_i\}_{i=0}^{n-1}$ , the net amount to buy, via futures contracts, at  $t_i$  (if  $u_i < 0$ , the agent sells  $|u_i|$  units), and  $\hat{u}$ , the amount of power to generate using thermal plants at  $T$ . Right before the transaction  $u_i$  at time  $t_i$ , three state variables are observable:  $x_i$ , the net position of future contracts;  $p_i$ , the market

price of electricity; and  $d_i$ , the residual demand. The dynamics of the states are given by  $x_{i+1} = x_i + u_i$ ,  $p_{i+1} = p_i + \nu u_i + \sigma_p \epsilon_i^{(1)} \sqrt{\Delta t}$ , and  $d_{i+1} = d_i + \mu \Delta t + \sigma_d \epsilon_i^{(2)} \sqrt{\Delta t}$  for  $i = 0, \dots, n-1$ . The price change is subject to a permanent linear price impact  $\nu u_i$  of the transactions and white noise  $\sigma_p \epsilon_i^{(1)} \sqrt{\Delta t}$ , where  $(\epsilon_i^{(1)}, \epsilon_i^{(2)})$  are independent vectors of binomial random variables each equal to 1 or  $-1$  with probability  $1/2$ . Denote the correlation of  $\epsilon^{(1)}$  and  $\epsilon^{(2)}$  by  $\rho$ . The random noise is a discrete-time approximation for the increment of a Brownian motion. The dynamics of the residual demand includes a drift term  $\mu \Delta t$  and white noise  $\sigma_d \epsilon_i^{(2)} \sqrt{\Delta t}$ . Note that  $\epsilon_i^{(1)}$  and  $\epsilon_i^{(2)}$  are unobserved at stage  $i$ , while the agent knows their distribution; see Online Appendix B. At time  $T$ , the agent can use his or her thermal power production with quantity  $\hat{u} \geq 0$  to match the target  $d_n - x_n$ .

The cost consists of three parts, i.e., the trading cost (profit) due to  $u_i$  for  $i = 0, \dots, n-1$ , in the form of  $\sum_{i=0}^{n-1} u_i(p_i + \gamma u_i)$ ; the production cost at time  $T$ ,  $\beta \hat{u}^2$ ; and the penalty for unmatched demand at time  $T$ ,  $\eta(d_n - x_n - \hat{u})^2$ . Here,  $\gamma \geq 0$  represents the linear temporary price impact. Given parameters  $(T, n, \gamma, \beta, \eta, \nu, \mu, \sigma_p, \sigma_d, \rho)$ , the agent faces the following optimization problem.<sup>11</sup>

$$\begin{aligned} \min_{\{u_i\}_{i=0}^{n-1}, \hat{u}} \quad & \mathbb{E} \left[ \sum_{i=0}^{n-1} u_i(p_i + \gamma u_i) + \beta \hat{u}^2 + \eta(d_n - x_n - \hat{u})^2 \right] \\ \text{s.t.} \quad & x_{i+1} = x_i + u_i \\ & p_{i+1} = p_i + \nu u_i + \sigma_p \epsilon_i^{(1)} \sqrt{\Delta t}, \\ & d_{i+1} = d_i + \mu \Delta t + \sigma_d \epsilon_i^{(2)} \sqrt{\Delta t}, \quad i = 0, \dots, n-1, \\ & \hat{u} \geq 0. \end{aligned}$$

**Remark 1.** The linear price impact can be alternatively modeled as a block-shaped LOB, of which the market depth is constant. More precisely, the temporary price impact  $\gamma$  corresponds to market depth  $1/(2\gamma)$ , which leads to the additional execution cost  $\gamma u^2$ . The difference between the permanent and temporary price impacts is caused by the resilience effect, i.e.,  $2\gamma u$  decays to  $\nu u$ . We demonstrate the different LOB models in this paper in Table 5.

As noted by Aid et al. (2016), the difficulty of an explicit solution lies in the constraint  $\hat{u} \geq 0$  at the terminal stage, which prohibits the use of the linear-quadratic regulator. However, the constraint can be easily incorporated into the class of MDPs discussed in Section 2. The formulation of the electricity management problem in the matrix form (2) is described in Online Appendix B.

For the pathwise convexity condition (4) to hold, after we substitute  $p_{i+1} = p_0 + \nu(u_0 + \dots + u_i)$  into the cost function, the quadratic coefficient matrix for  $\{u_i\}_{i=0}^{n-1}$  has  $\gamma$  on the diagonal and  $\nu/2$  off the diagonal<sup>12</sup>.

**Table 5.** Different Market Microstructure Represented by Order Books

Model	Market depth	Depth dynamics	Resilience	$m$	$l$
One LOB	$Q_i$	Markov chain	$e^{-\rho\Delta t}$	2	$\in \mathbb{Z}^+$
Multiple LOBs	$\{Q_{i,1}, \dots, Q_{i,N}\}$	Multivariate Markov chain	$\{e^{-\rho_{i,1}\Delta t}, \dots, e^{-\rho_{i,N}\Delta t}\}$	$N+1$	$\in \mathbb{Z}^+$
Linear impact	$1/(2\gamma)$	Constant	$v/(2\gamma)$	3	4

Assumption 1 and (4) hold if this matrix is positive definite, i.e.,  $\gamma > v/2 > 0$ , as the smallest eigenvalue equals  $\gamma - v/2$ . Note that the stagewise cost  $u_i(p_i + \gamma u_i)$  is not jointly convex in  $(u_i, x_i, p_i, d_i)$ , and thus, the algorithm in Birge and Louveaux (2011) cannot be applied.

For computation of the optimal policy and value function, note that the endogenous state space  $(x, p, d)$  can be reduced to two dimensions because only  $d - x$  needs to be tracked, instead of  $d$  and  $x$  separately. However, a constant state variable must be introduced for the partitioning algorithm to accommodate the drift term and the random noise, as in (5). Consequently,  $m$  equals 3 for the partitioning algorithm. The computation time of the partitioning algorithm and the grid-based scheme is listed in Table 2. The computation time of the grid-based scheme is acceptable at a 1% error level but not at a 0.1% error level. By comparison, the partitioning algorithm provides a useful benchmark to evaluate errors of approximation/discretization algorithms.

#### 4.2. Comparison to the Approximate Solution

We compare the exact optimal policy solved by the partitioning algorithm to the approximate solution in Aïd et al. (2016). The approximate solution relaxes the non-negativity constraint of  $\hat{u}$ ; as a result, the Markovian-jump linear quadratic regulator (Chizeck et al. 1986) can be used to analytically solve the unconstrained problem. The authors derive an error bound involving the probability that the constraint will be violated starting from an initial state, under the unconstrained linear-quadratic regulator. By contrast, our partitioning

algorithm solves the discrete-time constrained problem analytically. This method allows us to evaluate the optimal cost and compare it to the cost of the approximate solution. The comparison depends on the initial state  $(x_0, p_0, d_0)$ . For some initial states, the probability of  $\hat{u} \geq 0$  being active, i.e.,  $d_n - x_n \leq 0$ , is low under the optimal policy. Hence, the approximate solution is expected to work well for these initial states; for others, it may not.

Consider a numerical example whose parameters are given in the notes of Table 6. We do not consider high-frequency trading because transaction costs are high in electricity markets due to illiquidity (Schmalensee 2012). (Note that for  $n = 4$ , for example, the agent trades every six hours.) For each initial state  $(x_0, p_0, d_0)$ , the optimal cost  $J_{S_0,0}(x_0, p_0, d_0)$  is computed by the partitioning algorithm. The approximate policy is solved by MLQ (Chizeck et al. 1986). We simulate 10,000 realizations of random noise and take an average of the realized costs of the approximate solution.<sup>13</sup>

From Table 6, the performance of the approximate solution deteriorates noticeably as  $d_0 - x_0$  and  $p_0$  decrease. For small residual demand, i.e., small  $d_0 - x_0$ , the agent potentially faces the negative imbalance cost, i.e., holding a very large position of future contracts that leads to  $d_n - x_n \leq 0$  under random fluctuation. The agent should cautiously avoid such a scenario and target positive  $d_n - x_n$  by unwinding the position, as the positive imbalance cost can be effectively reduced by thermal power production, whose cost  $\beta$  is much less than the imbalance cost  $\eta$ . However, if the agent uses the approximate solution, he or she pretends that ther-

**Table 6.** Optimal Cost (OPT) and Cost of the Approximate Solution (APP) for Different Initial States and  $n$  and Their Percentage Differences

$(x_0, p_0, d_0)$	$n = 2$		$n = 3$		$n = 4$	
	OPT	APP	OPT	APP	OPT	APP
(0, 80, 20,000)	$8.5 \times 10^5$	$8.5 \times 10^5$ (0.6%)	$8.5 \times 10^5$	$8.5 \times 10^5$ (0.2%)	$8.5 \times 10^5$	$8.4 \times 10^5$ (-0.5%)
(0, 60, 15,000)	$5.0 \times 10^5$	$5.0 \times 10^5$ (0.9%)	$5.0 \times 10^5$	$5.0 \times 10^5$ (-0.3%)	$5.0 \times 10^5$	$5.0 \times 10^5$ (0.6%)
(0, 40, 10,000)	$2.5 \times 10^5$	$2.5 \times 10^5$ (0.6%)	$2.5 \times 10^5$	$2.5 \times 10^5$ (1.4%)	$2.5 \times 10^5$	$2.5 \times 10^5$ (0.6%)
(0, 30, 8,000)	$1.8 \times 10^5$	$1.8 \times 10^5$ (0.6%)	$1.8 \times 10^5$	$3.0 \times 10^6$ (> 100%)	$2.2 \times 10^5$	$2.0 \times 10^7$ (> 100%)
(0, 20, 6,000)	$1.8 \times 10^5$	$2.1 \times 10^7$ (> 100%)	$3.0 \times 10^5$	$7.6 \times 10^7$ (> 100%)	$3.1 \times 10^5$	$9.0 \times 10^7$ (> 100%)
(0, 10, 4,000)	$6.6 \times 10^5$	$2.1 \times 10^8$ (> 100%)	$6.6 \times 10^5$	$2.6 \times 10^8$ (> 100%)	$5.3 \times 10^5$	$2.3 \times 10^8$ (> 100%)

Notes. The approximate solution performs well in the first three initial states but poorly in the others. The realized costs of APP may be less than OPT (negative difference) because of the randomness of the simulation. The parameters are  $T = 24$  hours,  $\gamma = 0.2$ ,  $\beta = 0.002$ ,  $\eta = 100$ ,  $v = 4 \times 10^{-5}$ ,  $\mu = 0$ ,  $\sigma_p = 1$ ,  $\sigma_d = 1,000$ , and  $\rho = 0.8$ .

**Table 7.** Partition for  $i = 0$  and Optimal Policy  $u_0^*$  in Each Region with the Same Parameters as in Table 6 for  $n = 1$ 

Boundaries	Optimal policy	Probability
$-x_0 + 0.23p_0 + d_0 \leq -225,573$	$-0.98x_0 - 0.005p_0 + 0.98d_0$	1
$4,903 \geq -x_0 + 0.23p_0 + d_0 > -225,573$	$-0.96x_0 - 0.01p_0 + 0.96d_0 - 4,691$	0.5
$-x_0 + 0.23p_0 + d_0 > 4,903$	$-0.001x_0 - 0.23p_0 + 0.001d_0$	0

*Note.* The third column shows the probability of the constraint  $\hat{u} \geq 0$  being active at  $i = n$  under the optimal policy for initial states starting in each region (i.e., let  $d_1 = d_0 + \mu\Delta t + \sigma_d \epsilon_1^{(2)} \sqrt{\Delta t}$ ,  $x_1 = x_0 + u_0^*$ , and compute  $\mathbb{P}(d_1 - x_1 \leq 0)$ ).

mal plants can produce negative electricity  $\hat{u}$  and does not recognize the potential negative imbalance cost. Therefore, the relaxation would incur a high imbalance cost if  $d_n - x_n \leq 0$  indeed occurs. Similarly, for low price  $p_0$ , the agent tends to buy more from the intraday markets, and  $d_n - x_n$  is more likely to be negative. In both cases, the optimal policy tries to keep a higher level of  $d_i - x_i$  than the approximate solution to avoid a negative imbalance cost.

To illustrate the intuition, we list all three regions in the partition for  $n = 1$  and the same parameters as in Table 6. Table 7 shows that the probability of  $\hat{u}^* = 0$  under the optimal policy decreases in  $-x_0 - 0.23p_0 + d_0$ . As expected, in the last region, the approximate solution performs identically to the optimal policy computed by the partitioning algorithm because the constraint is active with zero probability. However, in the first two regions, the approximate solution can perform poorly, as shown by states four to six in Table 6. Moreover, for initial states in the first two regions, the optimal policy  $u_0^*$  is always greater than the approximate solution (equal to the optimal policy in the last region). This result implies that when the constraint is considered, the optimal policy buys more (or sells less) than the approximate solution to target positive  $d_n - x_n$ , thus confirming the aforementioned intuition. In summary, the analytical structure characterized by the partitioning algorithm gives us a better understanding of how the optimal policy is affected by the constraint.

## 5. Conclusion and Future Research

We study a class of Markov decision problems with linear state dynamics, quadratic cost, and linear inequality constraints. Based on a partition of the state space, the optimal policy and value function are linear and quadratic functions, respectively, of the endogenous state. We provide an algorithm that can compute the linear boundaries of the partition as well as the optimal policy and value function in each region. We conduct computational experiments to investigate how complexity (the number of regions in the partitions) grows in the problem dimension. As the results show, the partitions often have a moderate number of regions in the examples, much less than the theoretical upper bound.

There are two applications of the partition algorithm. First, we use the algorithm to study an LOB model with stochastic market depth, which captures various empirical phenomena. As shown in the numerical example, the optimal execution cost of a deterministic market depth model is significantly higher than that of our stochastic market depth model. We also apply the algorithm to study a renewable electricity management problem that previously had only an approximate solution. The partitioning algorithm allows us to derive the exact optimal policy, which considerably outperforms the approximate solution.

Methodologically, an interesting future direction would involve exploring whether the partitioning structure can be extended to more general classes of MDPs. For example, without much effort, one can incorporate piecewise quadratic and differentiable cost functions. A more difficult problem entails considering costs other than a quadratic function whose unconstrained minimizer can be easily solved. The key is to find a class of functions  $f(u, \mathbf{x})$  that are “closed” under the operation of minimization, i.e.,  $f(u^*(\mathbf{x}), \mathbf{x})$  is in the same class. One may also consider the case in which the transition probability matrix depends on the action; this leads to a similar difficulty, as  $P_{j,k} J_{i+1}(\mathbf{x}_{i+1}, s_k)$  in the Bellman equation is likely to have a higher order and is not included in the class of cost functions. Another possible extension is to allow multiple constraints in the same direction, e.g.,  $\max\{f'_1 \mathbf{x}_i, f'_2 \mathbf{x}_i\} \leq u_i$ . The value function is thus piecewise differentiable, and the solution is still subject to a partition of the state space, while quadratic boundaries may arise. For the complexity of the partitioning algorithm, the theoretical upper bound greatly overestimates the number of regions in a partition in many examples. Therefore, a more sophisticated complexity analysis might be needed to derive a tighter bound.

## Acknowledgments

The authors thank three anonymous reviewers and the associate editor for comments that greatly improved the paper.

## Endnotes

<sup>1</sup>This formulation is slightly different from the usual setting of random noise. The random noise at stage  $i$  is usually unknown at  $i$  but observed at  $i + 1$ . In this formulation we assume that  $S_i$  and the

coefficient matrices, including  $d_{s_i, i}$ , are realized and observed at  $i$ . However, the usual setting can be reduced to our formulation by introducing pre- and postdecision stages (Powell 2007). See Online Appendix B for more details on this transformation.

<sup>2</sup> Assumption 1 or the pathwise convexity condition need to hold analogously.

<sup>3</sup> Equation (1) can also be extended to similarly include constant terms and discrete additive noise by introducing the constant state.

<sup>4</sup> For a piecewise quadratic function, differentiability implies that the function differs by a multiple of  $x'LL'x$  across the boundary  $L'x = 0$ .

<sup>5</sup> To show this, note that there are three possibilities at every stage: The upper bound is binding, the lower bound is binding or neither. For a realized sample path of the Markov chain, there are potentially  $3^{n+1}$  regions at the initial stage. Because there are  $l^{n+1}$  such sample paths, a theoretical upper bound for the complexity is  $3^{(n+1)l^{n+1}}$ .

<sup>6</sup> See Online Appendix A for derivations of the bounds for both applications.

<sup>7</sup> In the LOB application,  $m = 2$  but  $n$  and  $l$  are adjustable. In the electricity application,  $m = 3$  and  $l = 4$  while  $n$  is adjustable. See Online Appendix B for details on the random generation of the parameters.

<sup>8</sup> Note that the pathwise convexity condition 4 is automatically satisfied for  $n = 10$  because  $s_1 > \exp(\rho\Delta t)s_l$ . While for  $n = 500$ , we impose a transition probability matrix such that  $Q_{i+1}$  can only transit to  $s_{j+1}$  or  $s_{j-1}$  if  $Q_i = s_j$ . Thus, the condition holds because  $s_j > \exp(\rho\Delta t)s_{j+1}$  for all  $j = 1, \dots, l-1$ , in which case  $l = 51$ .

<sup>9</sup> The grid-based scheme is also performed for comparison; see Table 2. We discretize  $x$  into 500 points and  $d$  into 100 to ensure that the discretization error is less than 0.1%.

<sup>10</sup> Similar to Proposition 2, the pathwise convexity (4) holds if  $Q_{i,j} > \exp(-2\rho_{i,j}\Delta t)Q_{i+1,j}$ .

<sup>11</sup> Aid et al. (2016) also investigate the case where the residual demand forecast is subject to sudden changes induced by prediction errors. In continuous time, it leads to a compound Poisson process in the dynamics of residual demand. In our discrete-time formulation, jumps can be naturally incorporated: We can add random noise  $\epsilon_i^j$  with distribution  $\mathbb{P}(\epsilon_i^j = 0) = 1 - \lambda\Delta t$  and  $\mathbb{P}(\epsilon_i^j = z_k) = f_k\lambda\Delta t$ , where  $z_k$  are possible jump sizes and  $f_k$  are their probabilities given that a jump occurs. We focus only on the case without jumps in this paper.

<sup>12</sup> The cost of the last stage is a sum of squares and convex. Thus, we only have to guarantee convexity for  $i = 0, \dots, n-1$ .

<sup>13</sup> The grid-based scheme is also performed for a comparison of computation time; see Table 2. We discretize  $d - x$  into 2,000 points and  $p$  into 1,000 to ensure that the discretization error is less than 0.1%.

## References

Ahn HJ, Bae KH, Chan K (2001) Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. *J. Finance* 56(2):767–788.

Aid R, Gruet P, Pham H (2016) An optimal trading problem in intraday electricity markets. *Math. Financial Econom.* 10(1):49–85.

Alfonsi A, Fruth A, Schied A (2010) Optimal execution strategies in limit order books with general shape functions. *Quant. Finance* 10(2):143–157.

Almgren R (2012) Optimal trading with stochastic liquidity and volatility. *SIAM J. Financial Math.* 3(1):163–181.

Almgren R, Chriss N (2001) Optimal execution of portfolio transactions. *J. Risk* 3(2):5–39.

Anupindi R, Morton TE, Pentico D (1996) The nonstationary stochastic lead-time inventory problem: Near-myopic bounds, heuristics, and testing. *Management Sci.* 42(1):124–129.

Athans M (1971) The role and use of the stochastic linear-quadratic-gaussian problem in control system design. *Automatic Control, IEEE Trans.* 16(6):529–552.

Bemporad A, Morari M, Dua V, Pistikopoulos EN (2002) The explicit linear quadratic regulator for constrained systems. *Automatica* 38(1):3–20.

Bertsekas DP (1995) *Dynamic Programming and Optimal Control*, Vol. 1 (Athena Scientific, Belmont, MA).

Bertsimas D, Lo AW (1998) Optimal control of execution costs. *J. Financial Markets* 1(1):1–50.

Biais B, Hillion P, Spatt C (1995) An empirical analysis of the limit order book and the order flow in the Paris Bourse. *J. Finance* 50(5):1655–1689.

Birge JR, Louveaux FV (2011) *Introduction to Stochastic Programming* (Springer, New York).

Blair Jr WPJ, Sworder DD (1975) Feedback control of a class of linear discrete systems with jump parameters and quadratic cost criteria. *Internat. J. Control* 21(5):833–841.

Bradley JR, Robinson LW (2005) Improved base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing Service Oper. Management* 7(4):319–329.

Brogaard J, Hendershott T, Riordan R (2014) High-frequency trading and price discovery. *Rev. Financial Stud.* 27(8):2267–2306.

Cao C, Hansch O, Wang X (2008) Order placement strategies in a pure limit order book market. *J. Financial Res.* 31(2):113–140.

Chen J, Feng L, Peng J, Ye Y (2014) Analytical results and efficient algorithm for optimal portfolio deleveraging with market impact. *Oper. Res.* 62(1):195–206.

Chiyachantana CN, Jain PK, Jiang C, Wood RA (2004) International evidence on institutional trading behavior and price impact. *J. Finance* 59(2):869–898.

Chizeck HJ, Willsky AS, Castanon D (1986) Discrete-time Markovian-jump linear quadratic optimal control. *Internat. J. Control* 43(1):213–231.

Cont R, Stoikov S, Talreja R (2010) A stochastic model for order book dynamics. *Oper. Res.* 58(3):549–563.

Ehrhardt R (1984) (s, S) policies for a dynamic inventory model with stochastic lead times. *Oper. Res.* 32(1):121–132.

Fodra P, Pham H (2015a) High frequency trading and asymptotics for small risk aversion in a Markov renewal model. *SIAM J. Financial Math.* 6(1):656–684.

Fodra P, Pham H (2015b) Semi-Markov model for market microstructure. *Appl. Math. Finance* 22(3):261–295.

Forsyth PA, Kennedy JS, Tse ST, Windcliff H (2012) Optimal trade execution: A mean quadratic variation approach. *J. Econom. Dynam. Control* 36(12):1971–1991.

Foucault T (1999) Order flow composition and trading costs in a dynamic limit order market. *J. Financial Markets* 2(2):99–134.

Foucault T, Kadan O, Kandel E (2005) Limit order book as a market for liquidity. *Rev. Financial Stud.* 18(4):1171–1217.

Fruth A, Schöneborn T, Urusov M (2014) Optimal trade execution and price manipulation in order books with time-varying liquidity. *Math. Finance* 24(4):651–695.

Glosten LR, Milgrom PR (1985) Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *J. Financial Econom.* 14(1):71–100.

Goettler RL, Parlour CA, Rajan U (2005) Equilibrium in a dynamic limit order market. *J. Finance* 60(5):2149–2192.

Goettler RL, Parlour CA, Rajan U (2009) Informed traders and limit order markets. *J. Financial Econom.* 93(1):67–87.

Gould MD, Porter MA, Williams S, McDonald M, Fenn DJ, Howison SD (2013) Limit order books. *Quant. Finance* 13(11):1709–1742.

Guilbaud F, Pham H (2013) Optimal high-frequency trading with limit and market orders. *Quant. Finance* 13(1):79–94.

Guo X, Zervos M (2015) Optimal execution with multiplicative price impact. *SIAM J. Financial Math.* 6(1):281–306.

Guo X, de Larrard A, Ruan Z (2013) Optimal placement in a limit order book. Working paper, University of California, Berkeley, Berkeley.

Henriot A (2014) Market design with centralized wind power management: Handling low-predictability in intraday markets. *Energy J.* 35(1):99–117.

- Horst U, Naujokat F (2014) When to cross the spread? Trading in two-sided limit order books. *SIAM J. Financial Math.* 5(1): 278–315.
- Huberman G, Stanzl W (2004) Price manipulation and quasi-arbitrage. *Econometrica* 72(4):1247–1275.
- Kavajecz KA (1999) A specialist's quoted depth and the limit order book. *J. Finance* 54(2):747–771.
- Kempf A, Korn O (1999) Market depth and order size. *J. Financial Markets* 2(1):29–48.
- Kyle AS (1985) Continuous auctions and insider trading. *Econometrica* 53(6):1315–1335.
- Louveaux FV (1980) A solution method for multistage stochastic programs with recourse with application to an energy investment problem. *Oper. Res.* 28(4):889–902.
- Moore JB, Zhou XY, Lim AE (1999) Discrete time lqg controls with control dependent noise. *Systems Control Lett.* 36(3): 199–206.
- Obizhaeva AA, Wang J (2013) Optimal trading strategy and supply/Demand dynamics. *J. Financial Markets* 16(1):1–32.
- O'Hara M (1995) *Market Microstructure Theory*, Vol. 108 (Blackwell, Cambridge, MA).
- Parlour CA, Seppi DJ (2008) Limit order markets: A survey. Thakor AV, Boot A, eds. *Handbook of Financial Intermediation and Banking*, Handbooks in Finance, Vol. 5 (Elsevier, Amsterdam), 63–96.
- Powell WB (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Vol. 703 (John Wiley & Sons, Hoboken, NJ).
- Predoiu S, Shaikhet G, Shreve S (2011) Optimal execution in a general one-sided limit-order book. *SIAM J. Financial Math.* 2(1):183–212.
- Puterman ML (2009) *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Vol. 414 (John Wiley & Sons, Hoboken, NJ).
- Ranaldo A (2004) Order aggressiveness in limit order book markets. *J. Financial Markets* 7(1):53–74.
- Roşu I (2009) A dynamic model of the limit order book. *Rev. Financial Stud.* 22(11):4601–4641.
- Saar G (2001) Price impact asymmetry of block trades: An institutional trading explanation. *Rev. Financial Stud.* 14(4):1153–1181.
- Schmalensee R (2012) Evaluating policies to increase electricity generation from renewable energy. *Rev. Environ. Econom. Policy* 6(1): 45–64.
- Tse ST, Forsyth PA, Kennedy JS, Windcliff H (2013) Comparison between the mean-variance optimal and the mean-quadratic-variation optimal trading strategies. *Appl. Math. Finance* 20(5): 415–449.
- Tsoukalas G, Wang J, Giesecke K (2016) Dynamic portfolio execution. Working paper, University of Pennsylvania, Philadelphia.
- Zipkin PH (2000) *Foundations of Inventory Management*, 1st ed. (McGraw-Hill, New York).