Efficient Simulation of Markov Chains Using Segmentation

Tim Brereton · Ole Stenzel · Björn Baumeier · Denis Andrienko · Volker Schmidt · Dirk Kroese

Abstract A methodology is proposed that is suitable for efficient simulation of continuous-time Markov chains that are nearly-completely decomposable. For such Markov chains the effort to adequately explore the state space via Crude Monte Carlo (CMC) simulation can be extremely large. The purpose of this paper is to provide a fast alternative to the standard CMC algorithm, which we call Aggregate Monte Carlo (AMC). The idea of the AMC algorithm is to reduce the jumping back and forth of the Markov chain in small subregions of the state space. We accomplish this by aggregating such problem regions into single states. We discuss two methods to identify collections of states where the Markov chain may become ‘trapped’: the stochastic watershed segmentation from image analysis, and a graph-theoretic decomposition method. As a motivating application, we consider the problem of estimating the charge carrier mobility of disordered organic semiconductors, which
contain low-energy regions in which the charge carrier can quickly become stuck. It is shown that the AMC estimator for the charge carrier mobility reduces computational costs by several orders of magnitude compared to the CMC estimator.

Keywords Markov chain · Nearly-completely decomposable · Monte Carlo · Segmentation · Watershed · Graph-theoretic decomposition · Electron transport · Mobility · Organic semiconductor

AMS 2000 Subject Classification 60J22

1 Introduction

Dynamic processes in complex physical systems often take place on multiple time scales, rendering a direct full evolution of the system, e.g., by solving a set of explicit deterministic equations, impossible. However, in infrequent-event systems of this type, i.e., systems with well-defined states and only occasional transitions between them, the dynamics can be treated in terms of the state-to-state transitions, described by a matrix of transition rates. An example of such a process is charge transport in disordered organic semiconductors. The dynamics of charges can be modeled by means of a continuous-time Markov chain (CTMC) on a graph, where the graph represents the transporting medium and the state of the process represents the positions of the charge carriers. The transition (or hopping) rates of the CTMC between adjacent vertices of the graph can be determined using charge transfer theories from the electronic and quantum mechanical properties of the semiconductor (see, for example, Rühle et al. 2011). One of the main characteristics to be measured is the charge carrier mobility \( \mu = v/|E| \), which (in the limit of a single mobile charge) depends on the drift velocity, \( v \), of the charge carrier as it passes through the random medium in an external electric field \( E \). The drift velocity \( v \) is the average velocity of the charge carrier in the direction of the electric field. This quantity can be estimated via Monte Carlo simulation of the CTMC—called kinetic Monte Carlo or dynamic Monte Carlo in the physics literature (see, e.g., Pasveer et al. 2005; Jansen 2012) Since the charge carrier mobility influences the performance of a material in technological applications, e.g., the efficiency of organic solar cells, simulation models for charge transport are a key ingredient of intensive efforts in in-silico design of high-efficiency organic semiconductors, see Baumeier et al. (2012).

The simulation approach introduced by Schönherr et al. (1981) and Bässler (1993), which we call crude Monte Carlo (CMC), has become a well-established method (see, e.g., Tessler et al. 2009; van der Holst et al. 2011). A major problem, however, is that for a large variety of materials the energy landscape associated with the random medium contains regions in which the charge carrier quickly becomes stuck. A consequence of this is that the estimation of drift velocity via CMC is not only very time-consuming (often to the extent of being practically infeasible) but also leads to unreliable estimators that can have a bias of several orders of magnitude. In addition, processes in organic electronics take place on multiple time and spatial scales. Therefore, it is essential to have computationally fast methods that analyze large system sizes over a long physical time without losing too much information about finer scale behavior.