Chapter 16
Multilevel Splitting

An alternative to importance sampling in estimating rare events and related functionals is *multilevel splitting*. In the context of estimating probabilities of a set $\mathscr{C}$ in path space, the multilevel splitting philosophy is to simulate particles that evolve according to the law of $\{X_i\}$, and at certain times split those particles considered more likely to lead to a trajectory that belongs to the set $\mathscr{C}$. For example, $\mathscr{C}$ might be the trajectories that reach some unlikely set B before hitting a likely set A, after starting in neither A nor B. In this case, the splitting will favor migration toward B. Splitting can also be used to enhance the sampling of regions that are important for a given integral. In all cases, particles which are split are given an appropriate weighting to ensure that the algorithm remains unbiased.

Broadly speaking, there are two types of multilevel splitting algorithms, those with killing and those without, where *stopping* is distinguished from killing. In the example just mentioned, particles are stopped upon entry into either A or B. Killing involves abandoning a particle prior to entry into either A or B, presumably because continuation of the trajectory is not worth the computational effort. Care must be taken that any killing will not introduce bias.

To the authors' knowledge, there is only one type of multilevel splitting algorithm without killing—the splitting algorithm (see [148] for further references). The standard implementation of this algorithm requires a sequence of sets $C_J \supset C_{J-1} \supset \cdots \supset C_0$, splitting thresholds called **splitting thresholds**, and a sequence of positive integers $R_{J-1}, \ldots, R_0$, splitting rates called **splitting rates**. A single particle is started at the initial position $x_0 \in C_J \backslash C_{J-1}$ and evolves according to the law of $\{X_i\}$. When a particle enters a set C_j for the first time, it produces $R_j - 1$ offspring. After splitting has occurred, all particles evolve independently of each other. Each particle is stopped according to whatever stopping rule is associated with $\{X_i\}$, and the algorithm terminates when all the particles generated have been stopped. The probability of interest is approximated by $N / \prod_{i=0}^{J-1} R_i$, where N is the number of particles simulated whose trajectories belong to $\mathscr{C}$. A more general version of this algorithm lets the splitting rates R_i take nonnegative real values, in which case the number of offspring is randomized.

A large deviation analysis of ordinary splitting is given in [76], which shows that it performs quite well when the thresholds are chosen properly. Although the splitting algorithm can be very effective, there is one clear source of inefficiency in dealing with rare events. The vast majority of the particles generated will not have trajectories that belong to the set $\mathscr{C}$, and so much of the computational effort is devoted to generating trajectories that do not make any direct contribution. Multilevel splitting algorithms with killing were introduced as a way to mitigate this problem. One of the first such algorithms was the RESTART (Repetitive Simulation Trials After Reaching Threshold) algorithm, introduced in [241, 242] (for others, see [208] and the references therein). Its implementation is identical to that of the standard splitting algorithm except that particles are split every time they enter a splitting threshold (and not just at the first entrance time, as with the standard splitting scheme), and particles are killed when they exit the splitting threshold in which they were born. The initial particle is assumed to be born in the set C_J, which by convention is equal to the state space of the process $\{X_i\}$, and so this particle is never killed.

The standard version of the RESTART algorithm requires that the splitting rates be integer-valued and that the process not cross more than one splitting threshold at each time step. However, its definition implicitly allows one to design an algorithm in which the process can cross more than one threshold in each time step. The issue of allowing the process to cross more than one threshold in any given time step was first addressed explicitly in the context of the DPR (Direct Probability Redistribution) algorithm, introduced in [152, 153].

In this chapter we will develop a theory for multilevel splitting, and in particular the RESTART/DPR algorithm, which parallels the theory for importance sampling that was developed in the last two chapters. Since the algorithm is notationally much more complicated than importance sampling, to simplify the presentation we consider only the case of estimating small probabilities, and refer to [77] for expected values.

Although the statements of performance analysis for splitting are often similar to those for importance sampling, it should be noted that there is an important distinction between the types of subsolution required for the two methods. For importance sampling, one needs functions that are *classical*-sense subsolutions. In contrast, splitting-type schemes require only a subsolution in a weak sense (see Definition 16.12). Indeed, this is in some sense expected, since importance sampling uses the gradient of the subsolution to construct the algorithm, while splitting uses only the function itself. For some problems it is easier to construct weak-sense subsolutions, in which case splitting-type schemes can be easier to apply. These and related issues will be discussed and illustrated by examples in Chap. 17.

When comparing importance sampling and splitting, one must recognize that the work used to produce samples need not be the same, and in fact, depending on circumstances, one method can be strongly favored over the other. However, when one uses subsolutions to design splitting schemes, the comparison simplifies somewhat, especially if the performance measure for importance sampling is the exponential decay rate. Suppose one were to consider, say, a work-normalized relative error [see (16.21)]. We will show that the computing cost of splitting grows

subexponentially when a subsolution is used. Thus the main issue in comparing splitting to importance sampling is to compare decay rates. This assumes that the implementation of importance sampling is relatively straightforward, i.e., given a subsolution, it is easy to compute the needed alternative sampling distribution. Since splitting simulates using only the original dynamics, it may be preferred when this is not the case, as in some multiscale models.

In the rest of this chapter, unless explicitly stated otherwise, by a splitting algorithm we mean the RESTART/DPR algorithm. We focus on this version of splitting, since in our experience it is usually preferable to ordinary splitting, and we will just note that analogous versions of all the statements presented here apply to ordinary splitting [76], and with much easier proofs. In Sect. 16.3, we derive formulae for the computational cost and second moment of the algorithm. These will be used in the asymptotic analysis, and Sect. 16.4 considers the asymptotic problem. In Sect. 16.4, a method for designing RESTART/DPR algorithms based on the subsolution framework is developed. Expressions for the asymptotic work-normalized error of such algorithms are derived using the formulas developed in the previous section, and subsolutions that lead to asymptotically optimal performance are identified. The formulation of splitting that is appropriate for finite-time problems of the same type as that considered in the context of importance sampling in Chaps. 14 and 15 is presented in Sect. 16.5.

16.1   Notation and Terminology

Let $\{X_i\}_{i \in \mathbb{N}_0}$ be a Markov chain with state space $\mathbb{R}^d$ for some $d \in \mathbb{N}$. Although we will later consider processes $\{X_i\}_{i \in \mathbb{N}_0}$ as elements of a sequence that satisfies a large deviation property, for notational simplicity the large deviation index is initially suppressed. Until Sect. 16.5, we focus on estimating

$$P_{x_0}\{X_M \in B\}, \tag{16.1}$$

where $M \doteq \inf\{i : X_i \in A \cup B\}$, and as in Sect. 14.1, A is open, B is closed, and $A \cap B = \emptyset$. Although not necessary, to simplify some arguments we will assume as in Remark 15.8 that $(A \cup B)^c$ is bounded and that its closure is denoted by D.

The following notation will be used. Branching processes, which take values in $\cup_{n=1}^{\infty}(\mathbb{R}^d)^n$, are denoted by $\{Z_i\}_{i \in \mathbb{N}_0}$. Each branching process has an $\mathbb{N}$-valued process $\{N_i\}_{i \in \mathbb{N}_0}$ associated with it, where N_i equals the number of particles present in the branching process at time i. As will be explained later in the section, particles are born through branching of existing particles when they reach certain thresholds, while particles die when they exit certain regions.

For each $i \in \mathbb{N}_0$ and $j = 1, \ldots, N_i$, $Z_{i,j}$ denotes the state of the jth particle at time i. We also define a measure on $\mathbb{R}^d$ associated with such a branching process by a random measure associated with branching processes

$$\bar{\delta}_{Z_i} \doteq \sum_{j=1}^{N_i} \delta_{Z_{i,j}}.$$

Note that this is typically not a probability measure, and it is referred to as the unnormalized empirical measure. If $Z_{i,j}$ is in either A or B, then it is killed at the next time step, and so will be counted only once in this measure. Note that this killing is distinct from the killing introduced for algorithmic efficiency.

Splitting schemes are often defined in terms of "importance functions." Later on, these importance functions will be identified with subsolutions translated by a constant, and we use U to denote such an object.[1] To be precise, an importance function is a continuous mapping $U : \mathbb{R}^d \to \mathbb{R}$ that is bounded from below. As we will see, it is only the relative values of $U(x)$ at different points that matter, and so we assume for simplicity of notation that $U(x) \geq 0$ for all $x \in \mathbb{R}^d$. There is also a parameter $\Delta \in (0, \infty)$ such that $R \doteq e^\Delta \in \{2, 3, \ldots\}$, and we define closed sets C_j by

$$C_j \doteq \{x \in D : U(x) \leq j\Delta\}$$

for $0 \leq j \leq J - 1 \doteq \lceil U(x_0)/\Delta \rceil - 1$ and $C_J \doteq D$. Note that $x_0 \notin C_{J-1}$. We also define a piecewise *constant* function $\bar{U}$ by setting $\bar{U}(x) = 0$ for $x \in C_0$ and $j\Delta$ if $x \in C_j \backslash C_{j-1}$

$$\bar{U}(x) = j\Delta \text{ for } x \in C_j \backslash C_{j-1}, \, j - 0, 1, \ldots, J,$$

where we follow the convention $C_{-1} = \emptyset$. After we introduce the large deviation scaling, it will be possible to obtain a collection of importance functions corresponding to a collection of values of the large deviation index from a single "generating" function in a convenient manner. While it would be possible to allow the splitting rate R or the spacing Δ between levels to depend on j, we will not do so once this scaling is used, and so to simplify notation, we have chosen not to do so here.

Later on, U will be derived from a subsolution $\bar{V}$ that satisfies a boundary condition (i.e., $\bar{V}(x) \leq 0$ for $x \in B$). One possibility will be to let $c \doteq \min[\bar{V}(x) : x \in B]$, with $c \leq 0$ due to the boundary condition, and then let $U(x)$ be equal to $[\bar{V}(x) - c] \vee 0$. In this case, as illustrated in Fig. 16.1, $C_0 \cap B$ may be smaller than B. Although the process stops when B is entered, if it crosses into C_1 or C_2 without entering B, the branching will continue. Thus the number of thresholds crossed before entering B depends on where B is entered. An alternative is to take $U(x)$ equal to $\bar{V}(x) \vee 0$, in which case $\bar{V}(x) \leq 0$ for $x \in B$ implies $B \subset C_0$, as in Fig. 16.2. The rate of decay of the second moment will be the same for both schemes, though one might expect slightly better performance from the first scheme.
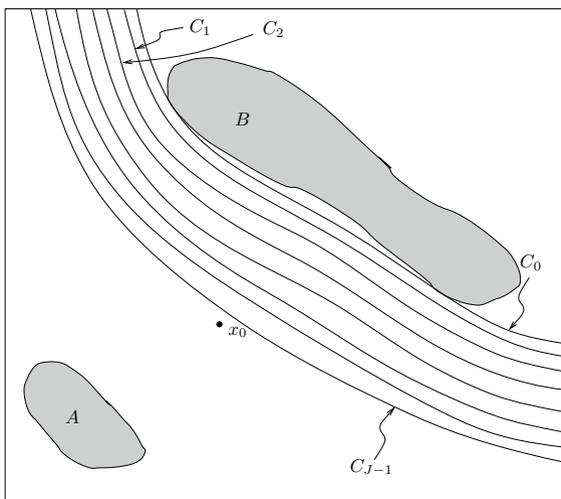
Given an importance function U and $x \in D$, let $\sigma(x)$ be the unique integer j such that $x \in C_j \backslash C_{j-1}$ unique integer j such that $x \in C_j \backslash C_{j-1}$. We let $\bar{U}_k$ denote

---

[1] While in the usual definition, importance functions *increase* as one approaches B, with the identification with subsolutions it will be convenient to have them decrease.
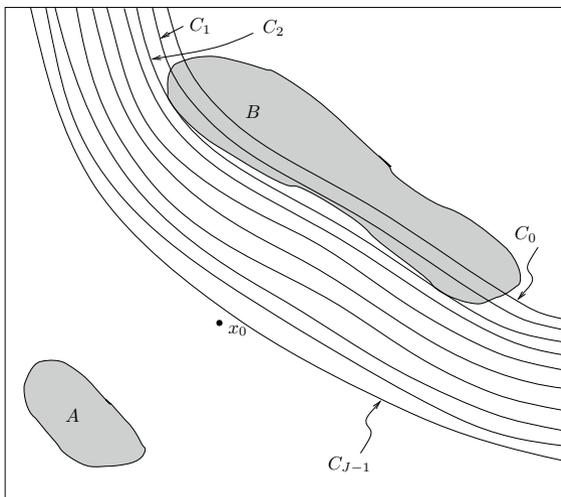
Fig. 16.1 Splitting based on $[\bar{V}(x) - c] \vee 0$

Fig. 16.2 Splitting based on $\bar{V}(x) \vee 0$

the common value of $\bar{U}(x)$ for all x such that $\sigma(x) = k$ $\bar{U}(x)$ if $\sigma(x) = k$ (i.e., $x \in C_k \backslash C_{k-1}$), and note that with this notation,

$$e^{-\bar{U}(x)} = e^{-\bar{U}_{\sigma(x)}} = e^{-\bar{U}_k} = R^{-k} \text{ for all } x \in C_k \backslash C_{k-1}, k = J, J-1, \ldots, 0. \quad (16.2)$$

With ordinary splitting, weights are assigned to particles so that a unit mass associated with a single particle starting at any location is partitioned evenly among the descendants at each splitting. Thus (16.2) are natural weightings for the given splitting rates, in that the fraction of the mass associated with each such descendant after k thresholds are crossed is R^{-k}. The issue is more subtle with RESTART, since particles reaching B can be the result of more splitting events than the number of intervening thresholds (due to multiple reentries of a particle into a splitting threshold). Nonetheless, owing to the killing, R^{-k} is still the correct weight to apply, as will be shown in the proof of unbiasedness.

In the standard version of the RESTART algorithm, splitting is fairly simple. Every time a particle enters a splitting threshold C_j, the deterministic number $R - 1$ of offspring are generated, so that including the parent, R particles result. Also, particles are destroyed when they exit the splitting threshold in which they are born, and thus each particle has an integer attached to it to record this splitting threshold. These are referred to as the support thresholds of the particles. While in general, one could allow the number of particles to be random and thereby accommodate arbitrary $R \in (0, \infty)$, there seems to be little practical benefit in doing so, and we restrict to the case in which R is an integer. In this chapter we will consider multilevel splitting, which accounts for the fact that the particles can jump more than one level in a single step. In such a case, different support thresholds must be assigned to the offspring. In order to analyze this mathematically, it is convenient to use the following notation. Let $\mathbb{S}$ be the set of elements $q \in \mathbb{N}_0^\infty$ such that $q_j = 0$ for all sufficiently large j. Vectors $q \in \mathbb{S}$ will be referred to as splitting vectors.

Consider a particle in a multilevel splitting algorithm that moves from $C_j \backslash C_{j-1}$ to $C_k \backslash C_{k-1}, k < j$, in a given time step. Then splitting will occur, and all offspring as well as the original parent particle will be located in $C_k \backslash C_{k-1}$. The support threshold of each new particle will be an element of $\{k, \ldots, j - 1\}$, and numbers of offspring and their support thresholds will be independent of all past data except through the values of j and k. It follows that the splitting of a particle is equivalent to assigning to each particle that splits a vector $q \in \mathbb{S}$. The number of *new* particles will be equal to $\sum_{l=0}^{\infty} q_l$, and precisely q_l of the new particles will be given support threshold l. Given that each particle generates $R - 1$ descendants upon moving from $C_{j+1} \backslash C_j$ to $C_j \backslash C_{j-1}$, it is clear that when moving from $C_j \backslash C_{j-1}$ to $C_k \backslash C_{k-1}$, we should use the splitting vector $q(j, k)$ defined by $q_l(j, k) \doteq 0$ if either $l \geq j$ or $l < k$, and splitting vectors

$$q_l(j, k) \doteq (R - 1)R^{j-l-1} \text{ if } k \leq j - 1 \text{ and } k \leq l \leq j - 1. \qquad (16.3)$$

Note that $\sum_{l \in \mathbb{N}_0} q_l(j, k) = R^{j-k} - 1$, and so including the original particle, exactly R^{j-k} particles are produced. We take $q_l(j, k) = 0$ for all l if $j \leq k$.

16.2   Formulation of the Algorithm

To define the algorithm, we assume the following condition. Conditions that imply the finiteness of M are given in Proposition 15.19.

Condition 16.1 *M is almost surely finite.*

Following the standard logic of acceleration methods generally, the hope is that with a well-chosen importance function $U(x)$, the variance of the estimator is made lower than that of standard Monte Carlo by building in information regarding the underlying process and the event of interest.

In order to analyze the algorithm, we will need some recursive formulas. Observe that if we examine a generic particle at some time after the algorithm has started, then it will be in a set of the form $C_j \backslash C_{j-1}$ and have a killing threshold in $\{j, \ldots, J\}$. For a Markov property to hold, if we imagine starting the process with an initial condition in $C_j \backslash C_{j-1}$, $j < J$, then the support threshold is part of the state variable, and so we must also assign a distribution to the support threshold that is consistent with the dynamics prior to entering $C_j \backslash C_{j-1}$. The distribution of the support threshold of the initial particle will be denoted by $\mathscr{I}$, and will be referred to as the *initializing distribution*. The correct form for such initializing distributions will be identified later on.

The estimator of (16.1), rewritten for a general initial condition x_0 (i.e., one not necessarily in $C_J \backslash C_{J-1}$), is

$$\sum_{i=0}^{\infty} \int_{\mathbb{R}^d} 1_B(x) e^{\bar{U}(x) - \bar{U}(x_0)} \bar{\delta}_{Z_i}(dx).$$

We recall that particles are killed the step after entering A or B, and so contribute to the sum for at most one time index. The weighting term $e^{\bar{U}(x) - \bar{U}(x_0)}$ is important when the number of thresholds crossed before reaching B depends on where the particle is located in B, as in Fig. 16.1.

The splitting thresholds, splitting rates, and splitting vectors of the algorithm will be defined using importance functions U and initialization distributions $\mathscr{I}$ as described previously. In Theorem 16.3 it will be shown that the algorithm is unbiased when the initializing distributions have a prescribed form that will be identified below. The algorithm, with the dependence on these quantities suppressed in the notation, can be written in pseudocode as follows.

RESTART/DPR Algorithm

```
Variables:
  i current time
  N_i number of particles at time i
  Z_{i,j} position of j^{th} particle at time i
  C_{i,j} current threshold of j^{th} particle at time i
  L_{i,j} support threshold of j^{th} particle at time i
  γ (at termination) an estimator of P_{x_0}{X_M ∈ B}
  j, k, l counting variables
  Y_{i,j} free variables
Initialization Step:
  N_0 = 1,  Z_{0,1} = x_0,  C_{0,1} = σ(x_0),  γ = 0,  i = 0
  generate a random variable L with distribution ℐ
  L_{0,1} = L
Main Algorithm:
  while N_i ≠ 0
    N_{i+1} = 0
    for  j = 1, ..., N_i
      Test to see if the particle is not killed due
      to stopping:
      if  Z_{i,j} ∉ A ∪ B
        generate a random variable Y_{i,j} with
        distribution  P{Y_{i,j} ∈ dy} = P{X_{i+1} ∈ dy | X_i = Z_{i,j}}

        Test to see if the particle is not killed
        due to threshold:
        if  σ(Y_{i,j}) ≤ L_{i,j}
          N_{i+1} = N_{i+1} + 1
          Z_{i+1,N_{i+1}} = Y_{i,j}
          C_{i+1,N_{i+1}} = σ(Y_{i,j})
          L_{i+1,N_{i+1}} = L_{i,j}
        end

        Test to see if particle should be branched:
        if  σ(Y_{i,j}) < C_{i,j}
          for  k = 1, ..., J
            for  l = 1, ..., q_k(C_{i,j}, σ(Y_{i,j}))
              N_{i+1} = N_{i+1} + 1
              Z_{i+1,N_{i+1}} = Y_{i,j}
              C_{i+1,N_{i+1}} = σ(Y_{i,j})
              L_{i+1,N_{i+1}} = k
            end
          end
        end
      end
    Test to see if the particle is stopped:
      if  Z_{i,j} ∈ A ∪ B
        γ = γ + 1_B(Z_{i,j}) e^{\bar{U}(Z_{i,j})}
      end
    end
    i = i + 1
  end
  γ = e^{-\bar{U}(x_0)} γ
```

Remark 16.2 The pseudocode just given presents a "parallel" version of the algorithm, in that all particles for a given threshold are split and then simulated either to the next threshold, the stopping criteria, or killing. Alternatively, one can implement a "sequential" version in which a particle is simulated until it either reaches the stopping criteria or is killed, recording where appropriate the number of additional particles that remain to be simulated for each threshold. After the current particle has been simulated to termination, the algorithm reverts to the highest threshold for which particles remain to be simulated, and starts a new particle.

Note that the output of the algorithm is indeed equal to the desired quantity

$$\gamma = e^{-\bar{U}(x_0)} \sum_{i=0}^{\infty} \int_{\mathbb{R}^d} 1_B(y) e^{\bar{U}(y)} \bar{\delta}_{Z_i}(dy). \tag{16.4}$$

An algorithm resulting from an importance function U, the collection of splitting vectors $q(j, k)$, and an initializing distribution $\mathscr{I}$ will be said to be unbiased if

$$E_{x_0} [\gamma] = P_{x_0}\{X_M \in B\}.$$

Recall that the splitting rates R are defined in terms the level $\Delta > 0$ that was used to partition U through $R = e^{\Delta}$. Define the vector

$$q_l \doteq \begin{cases} 1, & l = J, \\ (R - 1)R^{J-l-1}, & 0 \le l \le J - 1. \end{cases} \tag{16.5}$$

In the setting of ordinary splitting, where particles are branched only when they enter a threshold for the first time, q_l is the number of descendants that would be born in threshold l if all particles descending from a single particle in threshold J were to make it to l. We then define probability distributions λ_k on $\{k, \ldots, J\}$ by initializing distribution for splitting

$$\lambda_k(l) \doteq q_l/R^{J-k} = \begin{cases} R^{k-J}, & l = J, \\ (R - 1)R^{k-l-1}, & k \le l \le J - 1. \end{cases} \tag{16.6}$$

We extend the definition of λ_k to $\{0, \ldots, J\}$ by setting $\lambda_k(l) = 0$ for $l = 0, \ldots, k - 1$.

Theorem 16.3 *Fix $x_0 \in (A \cup B)^c$ and suppose that $X_0 = x_0$. Let U be an importance function, and assume Condition 16.1. If the initializing distribution is $\mathscr{I} = \lambda_{\sigma(x_0)}$ and the splitting vectors $q(j, k)$ are as in (16.3), then the resulting splitting algorithm is unbiased.*

The proof of Theorem 16.3 relies on the following lemma. Recall that as in the pseudocode, support threshold of particle m at time i

$$L_{i,m} \doteq \text{ support threshold of particle } m \text{ at time } i,$$

and that $\bar{\delta}_{Z_i} \doteq \sum_{m=1}^{N_i} \delta_{Z_{i,m}}$.

Lemma 16.4 *Assume Condition 16.1 and let h be a nonnegative function on $\mathbb{R}^d$. Let $\bar{h}(x) = h(x)e^{\bar{U}(x)}$ and let $i \in \mathbb{N}_0$ be given. For a splitting scheme with $\mathscr{I} = \lambda_{\sigma(x_0)}$ and $q(j,k)$ as in (16.3),*

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{m=1}^{N_i} \bar{h}(Z_{i,m}) 1_{\{L_{i,m}=l\}} \right] = E_{x_0} \left[h(X_i) \lambda_{\sigma(X_i)}(l) 1_{\{M \geq i\}} \right], \quad l = 0, 1, \ldots, J,$$

and

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\int_{\mathbb{R}^d} \bar{h}(y) \bar{\delta}_{Z_i}(dy) \right] = E_{x_0} \left[h(X_i) 1_{\{M \geq i\}} \right].$$

Proof We assume that $M > 0$, since otherwise, the lemma is trivial. The second result is obtained by summing the first one over l. Recall that $Z_{i,j}$ records the location of particle number j at time i. We will prove the first display by induction on i. The result holds for $i = 0$, since in this case there is only a single particle with support threshold distribution $\lambda_{\sigma(x_0)}(l)$, and thus

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\bar{h}(Z_{0,1}) 1_{\{L_{0,1}=l\}} \right] = e^{-\bar{U}(x_0)} \bar{h}(x_0) \lambda_{\sigma(x_0)}(l)$$
$$= E_{x_0} \left[h(X_0) \lambda_{\sigma(X_0)}(l) 1_{\{M \geq 0\}} \right].$$

Suppose the result has been proved up to some time i^*. We then claim that

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{j=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,j}) 1_{\{L_{i^*+1,j}=l\}} \right]$$
$$= e^{-\bar{U}(x_0)} E_{x_0} \left[e^{\bar{U}_{\sigma(Z_{0,1})} - \bar{U}_{\sigma(Z_{1,1})}} E_{Z_{1,1}} \left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m}) 1_{\{L_{i^*,m}=l\}} \right] \right]. \qquad (16.7)$$

Thus we currently have one particle located at x_0, whose support threshold lies in $\{\sigma(x_0), \ldots, J\}$. To prove the claim, we will compute the above expectation by conditioning on $Y_{0,1}$ as it appears in the pseudocode, which, we recall, has the distribution $P\{Y_{0,1} \in dy\} = P\{X_1 \in dy | X_0 = Z_{0,1}\}$. It suffices to show that for all $y \in \mathbb{R}^d$,

$$E_{x_0} \left[\sum_{j=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,j}) 1_{\{L_{i^*+1,j}=l\}} \middle| Y_{0,1} = y \right]$$
$$= e^{\bar{U}_{\sigma(x_0)} - \bar{U}_{\sigma(y)}} E_y \left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m}) 1_{\{L_{i^*,m}=l\}} \right]. \qquad (16.8)$$

Decomposing according to the support threshold, which has initializing distribution $\lambda_{\sigma(x_0)}(\cdot)$, we have

$$E_{x_0}\left[\sum_{j=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,j})1_{\{L_{i^*+1,j}=l\}} \;\middle|\; Y_{0,1} = y\right]$$

$$= \sum_{k=\sigma(x_0)}^{J} \lambda_{\sigma(x_0)}(k)E_{x_0,y,k}\left[\sum_{m=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,m})1_{\{L_{i^*+1,m}=l\}}\right], \qquad (16.9)$$

where $E_{x_0,y,k}$ denotes expected value given $Z_{0,1} = x_0$, $Y_{0,1} = y$, and $L_{0,1} = k$. Similarly, $E_{x,k}$ will denote the expected value given $Z_{0,1} = x$ and $L_{0,1} = k$. Note that by the Markov property,

$$E_{x_0,y,k}\left[\sum_{m=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,m})1_{\{L_{i^*+1,m}=l\}}\right]$$

$$= E_{x_0,y,k}\left[\sum_{r=1}^{N_1} E_{Z_{1,r},L_{1,r}}\left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m})1_{\{L_{i^*,m}=l\}}\right]\right].$$

Thus the expression in (16.9) can be written as

$$\sum_{k=\sigma(x_0)}^{J} \lambda_{\sigma(x_0)}(k)E_{x_0,y,k}\left[\sum_{r=1}^{N_1} E_{Z_{1,r},L_{1,r}}\left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m})1_{\{L_{i^*,m}=l\}}\right]\right]. \qquad (16.10)$$

We now consider the expression (16.10) for the three cases $\sigma(y) = \sigma(x_0)$, $\sigma(y) > \sigma(x_0)$, and $\sigma(y) < \sigma(x_0)$, and show that in each of these cases, the expression equals the right side of (16.8).

Consider the first case $\sigma(y) = \sigma(x_0)$. In this case, neither killing nor branching occurs, and so we have $N_1 = 1$, $Z_{1,1} = y$, and $L_{1,1} = L_{0,1} = k$. Thus (16.10) can be written as

$$\sum_{k=\sigma(y)}^{J} \lambda_{\sigma(y)}(k)E_{y,k}\left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m})1_{\{L_{i^*,m}=l\}}\right] = E_y\left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m})1_{\{L_{i^*,m}=l\}}\right],$$

which equals the right side of (16.8), since $e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}} = 1$.

Consider next the second case $\sigma(y) > \sigma(x_0)$. In this case, the particle has moved to a threshold with higher index, and branching does not occur. Recall that the particle is killed if and only if $k < \sigma(y)$, since this means that the particle exited its support threshold. Thus for $\sigma(x_0) \le k < \sigma(y)$, since $N_1 = 0$,

$$E_{x_0,y,k}\left[\sum_{r=1}^{N_1} E_{Z_{1,r},L_{1,r}}\left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m})1_{\{L_{i^*,m}=l\}}\right]\right] = 0.$$

Also, if $k \geq \sigma(y)$, then $N_1 = 1$ and $L_{1,1} = L_{0,1} = k$. Since $\lambda_{\sigma(x_0)}(k)/\lambda_{\sigma(y)}(k) = R^{\sigma(x_0)-\sigma(y)} = e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}}$, (16.10) in this case can be written as

$$
\sum_{k=\sigma(y)}^{J} \lambda_{\sigma(x_0)}(k) E_{y,k} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right]
$$

$$
= \sum_{k=\sigma(y)}^{J} \frac{\lambda_{\sigma(x_0)}(k)}{\lambda_{\sigma(y)}(k)} \lambda_{\sigma(y)}(k) E_{y,k} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right]
$$

$$
= e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}} \sum_{k=\sigma(y)}^{J} \lambda_{\sigma(y)}(k) E_{y,k} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right],
$$

which once more equals the right side of (16.8).

Finally, consider the case $\sigma(y) < \sigma(x_0)$. Here there is the possibility that new particles are created (i.e., $N_1 > 1$), though in all cases we have $Z_{1,r} = y$. When new particles are created, the associated thresholds are determined according to the measure $q_l(j, k)$, and so using (16.3) and the definition (16.6), (16.10) takes the form

$$
\sum_{k=\sigma(x_0)}^{J} \lambda_{\sigma(x_0)}(k) \left[\sum_{j=\sigma(y)}^{\sigma(x_0)-1} q_j(\sigma(x_0), \sigma(y)) E_{y,j} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right] \right.
$$

$$
\left. + E_{y,k} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right] \right].
$$

Since the sum $\sum_{j=\sigma(y)}^{\sigma(x_0)-1}$ has no k dependence, using $\sum_{k=\sigma(x_0)}^{J} \lambda_{\sigma(x_0)}(k) = 1$, that for $j \in \{\sigma(y), \ldots, \sigma(x_0) - 1\}$,

$$
q_j(\sigma(x_0), \sigma(y)) = e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}} \lambda_{\sigma(y)}(j),
$$

and that for $k \geq \sigma(x_0)$,

$$
\lambda_{\sigma(x_0)}(k) = e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}} \lambda_{\sigma(y)}(k),
$$

this quantity can be written as

$$
e^{\bar{U}_{\sigma(x_0)}-\bar{U}_{\sigma(y)}} \sum_{k=\sigma(y)}^{J} \lambda_{\sigma(y)}(k) E_{y,k} \left[\sum_{m=1}^{N_{i*}} \bar{h}(Z_{i*,m}) 1_{\{L_{i*,m}=l\}} \right].
$$

Thus in this case as well, (16.10) equals the right side of (16.8). This completes the proof of (16.8), and hence we have proved the claim in (16.7).

Thus from the induction hypothesis and (16.7), we have that

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{j=1}^{N_{i^*+1}} \bar{h}(Z_{i^*+1,j}) 1_{\{L_{i^*+1,j}=l\}} \right]$$

$$= e^{-\bar{U}(x_0)} E_{x_0} \left[e^{\bar{U}_{\sigma(Z_{0,1})} - \bar{U}_{\sigma(Z_{1,1})}} E_{Z_{1,1}} \left[\sum_{m=1}^{N_{i^*}} \bar{h}(Z_{i^*,m}) 1_{\{L_{i^*,m}=l\}} \right] \right]$$

$$= e^{-\bar{U}(x_0)} E_{x_0} \left[e^{\bar{U}_{\sigma(Z_{0,1})}} E_{Z_{1,1}} \left[h(X_{i^*}) \lambda_{\sigma(X_{i^*})}(l) 1_{\{M \geq i^*\}} \right] \right]$$

$$= E_{x_0} \left[E_{Z_{1,1}} \left[h(X_{i^*}) \lambda_{\sigma(X_{i^*})}(l) 1_{\{M \geq i^*\}} \right] \right]$$

$$= E_{x_0} \left[E_{X_1} \left[h(X_{i^*}) \lambda_{\sigma(X_{i^*})}(l) 1_{\{M \geq i^*\}} \right] \right]$$

$$= E_{x_0} \left[h(X_{i^*+1}) \lambda_{\sigma(X_{i^*+1})}(l) 1_{\{M \geq i^*+1\}} \right],$$

where the third equality uses the fact that $Z_{1,1}$ and X_1 have the same distribution, and the last equality uses the Markov property of $\{X_i\}$. This completes the induction step, and thus the lemma follows. $\qquad\square$

Proof (of Theorem 16.3) Since by Condition 16.1, $M < \infty$ a.s., we have

$$E_{x_0} [1_B(X_M)] = E_{x_0} \left[\sum_{i=0}^{\infty} 1_B(X_i) 1_{\{M=i\}} \right]$$

$$= \sum_{i=0}^{\infty} E_{x_0} \left[1_B(X_i) 1_{\{M=i\}} \right]$$

$$= \sum_{i=0}^{\infty} e^{-\bar{U}(x_0)} E_{x_0} \left[\int_{\mathbb{R}^d} 1_B(x) e^{\bar{U}(x)} \bar{\delta}_{Z_i}(dx) \right]$$

$$= E_{x_0} \left[\sum_{i=0}^{\infty} e^{-\bar{U}(x_0)} \int_{\mathbb{R}^d} 1_B(x) e^{\bar{U}(x)} \bar{\delta}_{Z_i}(dx) \right],$$

where the second and fourth equalities use Tonelli's theorem, and the third uses Lemma 16.4 applied to $h = 1_B$ and the observation that $1_B(X_i) 1_{\{M=i\}} = 1_B(X_i) 1_{\{M \geq i\}}$. The result now follows on observing that the term on the last line equals $E_{x_0} [\gamma]$. $\qquad\square$

16.3 Performance Measures

Recall that to derive a recurrence equation, we had to consider initializing distributions of the form $\mathscr{I} = \lambda_{\sigma(x_0)}$. In actual numerical implementation, it is always the case that $x_0 \in C_J$, which implies that all mass will be placed on $l = J$.

The performance of the algorithm depends on two factors: the second moment (and hence variance) of the estimator and the computational cost of each

simulation. To avoid discussion of any issues relating to the specific way the algorithm is implemented in practice, the computational cost is defined to be

$$w = \sum_{i=0}^{\infty} N_i, \tag{16.11}$$

where $N_i = \int_{\mathbb{R}^d} \bar{\delta}_{Z_i}(dx)$. Thus w is the sum of the lifetimes of all the particles simulated in the algorithm. In this section, formulas for both the second moment and computational cost are derived in terms of only the importance function and the underlying process. Throughout it is assumed that U is an importance function and Condition 16.1 is satisfied.

We begin by characterizing the mean of w.

Theorem 16.5 *Assume Condition 16.1. Then*

$$E_{x_0}[w] = e^{\bar{U}(x_0)} E_{x_0} \left[\sum_{i=0}^{M} e^{-\bar{U}(X_i)} \right].$$

Proof With the third equality due to Lemma 16.4 applied to $h(x) = e^{\bar{U}(x_0) - \bar{U}(x)}$, an application of Tonelli's theorem gives

$$\begin{aligned}
E_{x_0}[w] &= E_{x_0} \left[\sum_{i=0}^{\infty} \int_{\mathbb{R}^d} \bar{\delta}_{Z_i}(dx) \right] \\
&= \sum_{i=0}^{\infty} E_{x_0} \left[\int_{\mathbb{R}^d} \bar{\delta}_{Z_i}(dx) \right] \\
&= \sum_{i=0}^{\infty} e^{\bar{U}(x_0)} E_{x_0} \left[e^{-\bar{U}(X_i)} 1_{\{M \geq i\}} \right] \\
&= e^{\bar{U}(x_0)} E_{x_0} \left[\sum_{i=0}^{\infty} e^{-\bar{U}(X_i)} 1_{\{M \geq i\}} \right] \\
&= e^{\bar{U}(x_0)} E_{x_0} \left[\sum_{i=0}^{M} e^{-\bar{U}(X_i)} \right].
\end{aligned}$$

$\square$

Next note that the following bounds hold for all U, all $0 \leq k \leq l < j \leq J, 0 \leq k \leq m < j \leq J$. Since $q_l(j, k)$ as defined in (16.3) equals $[R^{j-l} - R^{j-l-1}]$, it follows that

$$\begin{aligned}
(q_l(j, k))^2 - q_l(j, k) &\leq [R^{j-l} - R^{j-l-1}]^2, \\
q_l(j, k) q_m(j, k) &= [R^{j-l} - R^{j-l-1}][R^{j-m} - R^{j-m-1}].
\end{aligned} \tag{16.12}$$

We can now give bounds for the second moment of the splitting estimator. Let $\mathfrak{S}(\bar{U})$ denote $E_{x_0}[\gamma^2]$ when $\bar{U}$ is used to design the splitting scheme.

Theorem 16.6 *Assume Condition 16.1. Then for all $x_0 \in (A \cup B)^c$,*

$$\mathfrak{S}(\bar{U}) \le e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{i=1}^{M} e^{\bar{U}(X_{i-1})} \left[P_{X_i}\{X_M \in B\} \right]^2 \right]$$
$$+ e^{-\bar{U}(x_0)} E_{x_0} \left[e^{\bar{U}(X_M)} 1_B(X_M) \right]. \tag{16.13}$$

Proof Recall that M is the first entry time of the set $A \cup B \subset D$. First consider the case in which there is $T < \infty$ such that $M \le T$ P_{x_0}-a.s. Let $W(x) \doteq e^{\bar{U}(x)} E_x[\gamma^2]$ with γ as in (16.4), and let $s(x, j; k)$, $k = 0, \ldots,$ denote iid sequences of random variables with the same distribution as γ, conditioned on $Z_{0,1} = x$ and $L_{0,1} = j$. Note that since the maximum possible number of particles is bounded, these random variables are bounded.

The proof is based on finding a recurrence equation for W. If $x_0 \notin A \cup B$, then there are two contributions to γ depending on the killing and/or splitting that takes place over the next time step. The first is due to future contributions if the particle stays within the support threshold, and the second occurs if new particles are generated $[\sigma(X_1) < \sigma(X_0)]$. To account for thresholds of both the existing particles and those that might be generated, let $Q_l(j, k)$ random vector defined in terms of $q_l(j, k)$ be random vectors defined by $Q_l(j, k) = q_l(j, k)$ for $j > l$ (i.e., these components are deterministic), and such that $Q_l(j, k)$ equals 1 for exactly one value of $j \le l \le J$ and 0 for remaining values, with the index chosen according to the initializing distribution λ_j. Recall that $q_l(j, k) = 0$ for all l if $k \ge j$. To abbreviate notation temporarily, let $\sigma_i = \sigma(Z_{i,1})$, $i = 0, 1$. Then

$$W(x_0) = e^{\bar{U}(x_0)} E_{x_0} \left[\left(1_{\{L_{0,1} \ge \sigma_1\}} e^{\bar{U}(Z_{1,1}) - \bar{U}(Z_{0,1})} s(Z_{1,1}, L_{0,1}; 0) \right. \right.$$

$$\left. \left. + 1_{\{\sigma_0 > \sigma_1\}} \left(\sum_{j=\sigma_1}^{\sigma_0-1} \sum_{m=1}^{q_j(\sigma_0,\sigma_1)} e^{\bar{U}(Z_{1,1}) - \bar{U}(Z_{0,1})} s(Z_{1,1}, j; m) \right) \right)^2 \right]$$

$$= e^{\bar{U}(x_0)} E_{x_0} \left[\left(\sum_{j=0}^{J} \sum_{m=1}^{Q_j(\sigma_0,\sigma_1)} e^{\bar{U}(Z_{1,1}) - \bar{U}(Z_{0,1})} s(Z_{1,1}, j; m) \right)^2 \right].$$

We now use the following facts: $L_{0,1}$ has distribution $\lambda_{\sigma(X_0)}$; $Z_{1,1}$ has the same distribution (conditioned on $Z_{0,1} = X_0 = x_0$) as X_1; by the definition of $Q_l(j, k)$, for all j, k, l [see also (16.6) and (16.5)],

$$E_{x_0} Q_l(j, k) e^{\bar{U}_k - \bar{U}_j} = \lambda_k(l); \tag{16.14}$$

and that the future evolution of the algorithm is independent of the $Q_l(j, k)$. Now let σ_i denote $\sigma(X_i)$, $i = 0, 1$. Together with the expression just given for $W(x_0)$, these give

$$W(x_0) = e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{j,k=0}^{J} e^{2\bar{U}(X_1)} Q_j(\sigma_0, \sigma_1) Q_k(\sigma_0, \sigma_1) E_{X_1, j}[\gamma] E_{X_1, k}[\gamma] \right]$$

(16.15)

$$+ e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{j=0}^{J} e^{2\bar{U}(X_1)} Q_j(\sigma_0, \sigma_1) \left(E_{X_1, j}[\gamma^2] - \left(E_{X_1, j}[\gamma] \right)^2 \right) \right].$$

We examine the various terms separately. Using (16.14) and $W(x) \doteq e^{\bar{U}(x)} E_x[\gamma^2]$,

$$e^{-\bar{U}(x_0)} E_{x_0} \left[\sum_{j=0}^{J} e^{2\bar{U}(X_1)} Q_j(\sigma_0, \sigma_1) E_{X_1, j}[\gamma^2] \right]$$

$$= E_{x_0} \left[e^{\bar{U}(X_1)} \sum_{j=0}^{J} e^{\bar{U}(X_1) - \bar{U}(X_0)} Q_j(\sigma_0, \sigma_1) E_{X_1, j}[\gamma^2] \right]$$

$$= E_{x_0} \left[e^{\bar{U}(X_1)} \sum_{j=0}^{J} \lambda_{\sigma_1}(j) E_{X_1, j}[\gamma^2] \right]$$

$$= E_{x_0}[W(X_1)].$$

(16.16)

If (16.16) is subtracted from the right side of (16.15), the remaining quantity is

$$e^{\bar{U}(x_0)} E_{x_0} \left[\sum_{j=0}^{J} \sum_{l=0}^{J} e^{2\bar{U}(X_1) - 2\bar{U}(X_0)} Q_j(\sigma_0, \sigma_1) Q_l(\sigma_0, \sigma_1) E_{X_1, j}[\gamma] E_{X_1, l}[\gamma] \right]$$

$$- e^{\bar{U}(x_0)} E_{x_0} \left[\sum_{j=0}^{J} e^{2\bar{U}(X_1) - 2\bar{U}(X_0)} Q_j(\sigma_0, \sigma_1) \left(E_{X_1, j}[\gamma] \right)^2 \right].$$

(16.17)

The terms with both l and j at or above $\sigma(X_0)$ contribute nothing to this expression. Indeed, $Q_j(\sigma_0, \sigma_1)$ is equal to 1 for exactly one j and to 0 for all remaining j that are greater than $\sigma(X_0) - 1$. Hence the corresponding terms in the double and single sums cancel. Also, this is the only possibility when $\sigma(X_1) \geq \sigma(X_0)$, and so we restrict to $\sigma(X_1) < \sigma(X_0)$, and use that $Q_j(\sigma_0, \sigma_1) = 0$ for $j < \sigma(X_1)$ when this is the case. Dropping terms that contribute nothing, we decompose the double sum as

$$\sum_{j=\sigma(X_1)}^{\sigma(X_0)-1} \sum_{l=\sigma(X_1)}^{\sigma(X_0)-1} +2 \sum_{j=\sigma(X_0)}^{J} \sum_{l=\sigma(X_1)}^{\sigma(X_0)-1}.$$

Using (16.14), we get the following upper bound for expression (16.17):

$$e^{\bar{U}(x_0)} E_{x_0}\left[1_{\{\sigma(X_1)<\sigma(X_0)\}}\left(\sum_{j=\sigma(X_1)}^{\sigma(X_0)-1} \lambda_{\sigma(X_1)}(j)E_{X_1,j}\left[\gamma\right]\right)^2\right]$$

$$+ 2e^{\bar{U}(x_0)} E_{x_0}\left[1_{\{\sigma(X_1)<\sigma(X_0)\}}\left(\sum_{j=\sigma(X_0)}^{J} \lambda_{\sigma(X_1)}(j)E_{X_1,j}\left[\gamma\right]\right)\right.$$

$$\left. \times \left(\sum_{l=\sigma(X_1)}^{\sigma(X_0)-1} \lambda_{\sigma(X_1)}(l)E_{X_1,l}\left[\gamma\right]\right)\right]$$

$$\leq e^{\bar{U}(x_0)} E_{x_0}\left[1_{\{\sigma(X_1)<\sigma(X_0)\}}\left(\sum_{j=\sigma(X_1)}^{J} \lambda_{\sigma(X_1)}(j)E_{X_1,j}\left[\gamma\right]\right)^2\right]. \qquad (16.18)$$

We now combine (16.15), (16.16), (16.18), and Theorem 16.3 to get that for $x_0 \notin A \cup B$,

$$W(x_0) \leq e^{\bar{U}(x_0)} E_{x_0}\left[1_{\{\sigma(X_1)<\sigma(X_0)\}}\left(E_{X_1}\left[1_B(X_M)\right]\right)^2\right] + E_{x_0}\left[W(X_1)\right].$$

Since all functions involved are bounded and nonnegative, it follows that the sequence

$$\Sigma_i \doteq W(X_{i\wedge M}) + \sum_{j=1}^{i\wedge M}\left\{e^{\bar{U}(X_{j-1})}\left(E_{X_j}\left[1_B(X_M)\right]\right)^2\right\}$$

defined for $i \in \{0, \dots, T\}$ is a submartingale. Thus, using $W(X_{T\wedge M}) = e^{\bar{U}(X_M)}$ $1_B(X_M)$ and that $1_B(X_k) = 0$ if $k < M$, we have

$$e^{\bar{U}(x_0)}\mathfrak{S}(\bar{U}) = W(x_0)$$
$$= \Sigma_0$$
$$\leq E_{x_0}\left[\Sigma_T\right]$$
$$= E_{x_0}\left[\sum_{i=1}^{M} e^{\bar{U}(X_{i-1})}\left(E_{X_i}\left[1_B(X_M)\right]\right)^2\right] + E_{x_0}\left[e^{\bar{U}(X_M)}1_B(X_M)\right],$$

which is the same as (16.13).

We next remove the restriction that $M \leq T$ for some constant $T < \infty$. We add time as a state variable [i.e., work with the process (X_i, i)], and consider the analogous

estimation problem in which the stopping set is $(A \cup B) \times \{T\}$ (i.e., we stop if either X_i enters $A \cup B$ or $i = T$). Then γ_T defined in an analogous manner is an unbiased estimator of $E_{x_0}\left[1_B(X_M)1_{\{M \leq T\}}\right]$, and by the previous result for bounded stopping times,

$$e^{\bar{U}(x_0)}E_{x_0}[(\gamma_T)^2] \tag{16.19}$$
$$\leq E_{x_0}\left[\sum_{i=1}^{M \wedge T} e^{\bar{U}(X_{i-1})}\left(E_{X_i}\left[1_B(X_M)1_{\{M \leq T\}}\right]\right)^2\right] + E_{x_0}\left[e^{\bar{U}(X_M)}1_B(X_M)1_{\{M \leq T\}}\right].$$

Also note that

$$\gamma_T = e^{-\bar{U}(x_0)}\sum_{i=0}^{T}\int_{\mathbb{R}^d} e^{\bar{U}(y)}1_B(y)\bar{\delta}_{Z_i}(dy)$$

and $\gamma_T \uparrow \gamma$ a.s. By the monotone convergence theorem,

$$E_{x_0}\left[(\gamma_T)^2\right] \to \mathfrak{S}(\bar{U}).$$

Using this in (16.19), the nonnegativity of 1_B, and the monotone convergence theorem a second time gives (16.13) without the restriction on M. $\qquad\square$

The following result gives a lower bound on the second moment of the estimator, complementing the upper bound in Theorem 16.6.

Theorem 16.7 *Assume Condition 16.1. Then*

$$\mathfrak{S}(\bar{U}) \geq e^{-\bar{U}(x_0)}E_{x_0}\left[e^{\bar{U}(X_M)}1_B(X_M)\right].$$

Proof From the nonnegativity of 1_B, (16.15), and (16.16), it follows that $W(x_0) \geq E_{x_0}[W(X_1)]$. From the Markov property of $\{X_i\}$, it follows that $\Sigma_i \doteq W(X_{i \wedge M})$ is a supermartingale, and in particular,

$$E_{x_0}[W(X_{M \wedge i})] \leq W(x_0).$$

The definition $W \doteq e^{\bar{U}(x)}E_x[\gamma^2]$ and its nonnegativity then give

$$E_{x_0}\left[W(X_M)1_{\{M \leq i\}}\right] \leq E_{x_0}[W(X_{M \wedge i})] \leq e^{\bar{U}(x_0)}\mathfrak{S}(\bar{U}).$$

Since $W(x) = e^{\bar{U}(x)}1_B(x)$ for $x \in A \cup B$, the last display gives

$$E_{x_0}\left[e^{\bar{U}(X_M)}1_B(X_M)1_{\{M \leq i\}}\right] \leq e^{\bar{U}(x_0)}\mathfrak{S}(\bar{U}).$$

The result now follows on sending $i \to \infty$ and using the monotone convergence theorem. $\qquad\square$

16.4   Design and Asymptotic Analysis of Splitting Schemes

Thus far we have considered only the problem of estimating a single probability of the form (16.1). Now we shall turn to the problem of estimating a sequence of such probabilities

$$P_{x_0}\{X_{M^n}^n \in B\}, \tag{16.20}$$

$n \in \mathbb{N}$, where $M^n \doteq \inf\{i : X_i^n \in A \cup B\}$ and $\{X_i^n\}_{i \in \mathbb{N}_0}$ is a Markov chain for each n that satisfies a large deviation principle as $n \to \infty$ (see Condition 16.9). We recall that by assumption, A is open, B is closed, and $(A \cup B)^c$ is bounded. With $\mathfrak{S}^n(\bar{U})$ denoting the second moment $E_{x_0}\left[(\gamma^n)^2\right]$, the asymptotic performance will be evaluated using the following measure of work-normalized error:

$$\lim_{n \to \infty} \frac{1}{n} \log \frac{\mathfrak{S}^n(\bar{U}) E_{x_0}[w^n]}{\left[P_{x_0}\{X_{M^n}^n \in B\}\right]^2}, \tag{16.21}$$

where γ^n is the splitting-based estimator for (16.20), and w^n is its computational cost, which was defined in the nonasymptotic setting in (16.11). (Such a weighted performance measure is not needed for importance sampling, since the cost per sample is essentially independent of the subsolution.)

Suppose that $-(1/n) \log P_{x_0}\{X_{M^n}^n \in B\} \to V(x_0)$ as $n \to \infty$. Jensen's inequality as discussed in Chap. 14 shows that the best possible value of (16.21) is zero, and this occurs only when the work grows subexponentially and the second moment $\mathfrak{S}^n(\bar{U})$ decays at rate $2V(x_0)$. Bounds on the asymptotic behavior of the work-normalized error will be obtained using Theorems 16.5–16.7 and are stated in Theorem 16.15, Corollary 16.16, and Theorem 16.18.

Remark 16.8 As in Chap. 15, the theoretical bounds on performance are given for the case of a fixed initial condition x_0. However, all results are easily generalized to the case of varying initial conditions x_n that converge to x_0 as $n \to \infty$. This generalization is useful for systems with discrete state spaces, such as queueing networks.

The theory presented in this section will require some fairly standard assumptions on the stability and large deviation behavior of $\{X_i^n\}$, and also some regularity properties on A and B that are qualitatively similar to assumptions made in Chap. 15 (e.g., Condition 15.9). For example, we will want to know that $\tau^n \doteq M^n/n$ can essentially be taken as bounded, in the sense that there is some $T < \infty$ such that the event $\tau^n > T$ is unimportant as far as the large deviation asymptotics are concerned. This is an important qualitative assumption, and it is related to stability properties of the law of large numbers limit processes obtained when $n \to \infty$.

We define continuous time stochastic processes as usual by setting $X^n(t) = X_i^n$ for $t = i/n$ and by piecewise linear interpolation for $t \in [i/n, (i+1)/n)$. Throughout, we assume that $x_0 \in (\bar{A} \cup B)^c$. The following condition will be needed to establish

a limit for the second moment; it is not needed if one wants to give just upper bounds on the second moment.

Condition 16.9 *For every $T \in (0, \infty)$, the sequence $\{X^n\}_{n \in \mathbb{N}}$ satisfies a large deviation principle on $\mathscr{C}\left([0, T] : \mathbb{R}^d\right)$ that is uniform with respect to the initial condition in compacts sets. The rate function is of the form*

$$I_T(\phi) \doteq \int_0^T L(\phi(s), \dot{\phi}(s)) ds$$

if $\phi \in \mathscr{C}\left([0, T] : \mathbb{R}^d\right)$ is absolutely continuous with $\phi(0) = x_0$ and ∞ otherwise, where L is a nonnegative measurable function.

As remarked above, the conditions we use beyond the LDP can be partitioned into "stability" and "controllability" type conditions. We give two conditions that will be sufficient (but not necessary) for what follows. Moreover, the sufficient conditions we give will by themselves cover many interesting problems. The stability condition (Condition 16.10) will imply that the algorithm is practical in that the tails of the hitting times are controlled, and also that the escape time problem can be approximated using estimates over finite time intervals. The condition we refer to as "controllability" (Condition 16.11) is needed to establish limits rather than just bounds, and is analogous to the additional conditions that would have been required in Chap. 15 as noted in Remark 15.16.

We will assume the following condition, which is the same as Condition 15.9 in Chap. 15.

Condition 16.10 *There exist $c > 0$, $T_0 \in (0, \infty)$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $T < \infty$, and $x \in D$,*

$$P_x\{\tau^n > T\} \leq \exp\{-cn(T - T_0)\}.$$

Note that Condition 16.10 implies that

$$\limsup_{T \to \infty} \limsup_{n \to \infty} \sup_{x \in D} \frac{1}{n} \log P_x\{\tau^n > T\} = -\infty. \tag{16.22}$$

Condition 16.10 would not hold if there were two attractors for the zero-cost trajectories, $A \cup B$ contains one of the attractors but not the other, and the process starts in the domain of attraction of the stable point that is not in $A \cup B$.

Condition 16.11 *Suppose we are given absolutely continuous ϕ satisfying $\phi(0) = x_0 \notin \bar{A} \cup B$, $\phi(t) \notin A \cup B^\circ$ for $t \in [0, T)$, and $\phi(T) \in B$ for some $T < \infty$. Then given $\gamma > 0$, there exist absolutely continuous ϕ^*, $T^* < \infty$, and $\tau^* < T^*$ such that $\phi^*(0) = x_0$, $\phi^*(t) \notin \bar{A} \cup B$ for $t \in [0, \tau^*)$, $\phi^*(t) \in B^\circ$ for $t \in (\tau^*, T^*]$, and such that*

$$\int_0^{T^*} L(\phi^*(r), \dot{\phi}^*(r))dr \le \int_0^T L(\phi(r), \dot{\phi}(r))dr + \gamma, \ \left\| \phi(T) - \phi^*(\tau^*) \right\| \le \gamma.$$

One can consider this a controllability-type condition. It says that given a trajectory ϕ that enters $\bar{A} \cup B$ but not $A \cup B^\circ$ and finally enters B° at T, one can find a trajectory with almost the same cost that avoids $\bar{A} \cup B$ till τ^*, at which time it enters B° near $\phi(T)$. One can establish more concrete conditions that imply this condition, such as assuming that $L(x, \beta)$ is continuous, bounded on each compact subset of $\mathbb{R}^d \times \mathbb{R}^d$, and assuming regularity properties for the boundaries of A and B.

We next give a definition of subsolution appropriate to this problem, but phrased directly in terms of the calculus of variations problem. The definition via calculus of variations is somewhat more to the point of what is required and is used in the proofs. For $y \in (A \cup B^\circ)^c$ and $T \in (0, \infty)$, define

$$K_{y,T} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (16.23)$$
$$\doteq \left\{ \phi \in \mathscr{AC}([0, T] : \mathbb{R}^d) : \phi(0) = y; \phi(s) \notin A \cup B^\circ, s \in (0, T), \phi(T) \in B \right\}.$$

Definition 16.12 A continuous function $\bar{V} : \mathbb{R}^d \to \mathbb{R}$ is a subsolution if it is bounded from below,

$$\bar{V}(y) \le \inf_{\phi \in K_{y,T}, T < \infty} \left[\int_0^T L(\phi(s), \dot{\phi}(s))ds + \bar{V}(\phi(T)) \right] \qquad (16.24)$$

for all $y \in (A \cup B^\circ)^c$, and $\bar{V}(z) \le 0$ for $z \in B$.

Remark 16.13 (*Relations between notions of subsolution I*) Suppose that $\bar{V}$ is a subsolution in the sense of Definitions 14.4 or 14.5 that is bounded from below, and to simplify the discussion, assume also that $\mathbb{H}(x, \alpha)$ is continuous. Then we claim that $\bar{V}$ is a subsolution in the sense of Definition 16.12. Consider the case of Definition 14.4, and for $y \in (A \cup B^\circ)^c$, suppose $\phi \in K_{y,T}$. Since $\bar{V}$ is continuously differentiable, the definition $\mathbb{H}(x, \alpha) \doteq \inf_{\beta \in \mathbb{R}^d} [\langle \alpha, \beta \rangle + L(x, \beta)]$ implies

$$\left\langle D\bar{V}(\phi(s)), \dot{\phi}(s) \right\rangle + L(\phi(s), \dot{\phi}(s)) \ge 0$$

for a.e. $s \in [0, T]$. Integrating gives

$$\bar{V}(y) \le \left[\int_0^T L(\phi(s), \dot{\phi}(s))ds + \bar{V}(\phi(T)) \right].$$

Since $\phi \in K_{y,T}$ is arbitrary, and $\bar{V}(z) \le 0$ for $z \in B$ is part of Definition 14.4, the claim follows. For the case of piecewise classical subsolutions it is enough to note that the mollification (14.16) produces a classical-sense subsolution $\bar{V}^\delta$, and the claim follows by taking the limit $\delta \downarrow 0$. Note that one does not need to use the mollified subsolution for the design of the splitting scheme, but can instead use the potentially

nonsmooth limit. It is also worth noting that the continuity of $\mathbb{H}$ is not used in any essential way, so that the analogous claim holds for problems involving discontinuous statistics, such as queueing networks.

Remark 16.14 (*Relations between notions of subsolution II*) The set of trajectories $K_{y,T}$ in (16.23) differs from $C_{y,T}$ introduced in Chap. 14 in that trajectories are excluded from B for $s \in (0, T)$ for $C_{y,T}$, but only from B° for $K_{y,T}$. The reason for the difference is the slightly different way in which the subsolution property is used in the cases of importance sampling and splitting. However, under the conditions we assume, to obtain a limit as in Theorem 16.15 one could use $C_{y,T}$ in Definition 16.12 instead. Indeed, since $C_{y,T} \subset K_{y,T}$, one has only to check that the defining inequality (16.24) holds for $\phi \in K_{y,T}$ for all T if it holds for all $\phi \in C_{y,S}$, $y \in (A \cup B^\circ)^c$, and $S \in (0, \infty)$. But this is easy to check under Condition 16.11. Note also that the two definitions are equivalent without reference to Condition 16.11 if $\bar{V}$ is constant on B, simply because $L \geq 0$.

Note that $\bar{V}$ is never greater than the solution to the calculus of variations problem, which is defined by $V(z) = 0$ for $z \in B$, $V(z) = \infty$ for $z \in A$, and

$$V(x) = \inf_{\phi \in K_{x,T}; T < \infty} \left[\int_0^T L(\phi(s), \dot{\phi}(s)) ds \right] \qquad (16.25)$$

for $x \in (A \cup B)^c$. Given a subsolution, the corresponding splitting scheme is defined as follows. Thresholds are defined in terms of the levels

$$\bar{V}(x_0), \bar{V}(x_0) - (\log R)/n, \bar{V}(x_0) - (2 \log R)/n, \ldots.$$

Let J^n be the smallest number such that $\bar{V}(x) \geq \bar{V}(x_0) - (J^n \log R)/n$ for all $x \in D$, so that there are no more than J^n thresholds. Then we define $C_{J^n}^n \doteq D, C_{-1}^n \doteq \emptyset$, and

$$C_{J^n - j}^n \doteq \left\{ x : \bar{V}(x) \leq \bar{V}(x_0) - (j \log R)/n \right\}, \quad j = 1, \ldots, J^n. \qquad (16.26)$$

Recall $R = e^\Delta$ and define $\Delta^n \doteq \Delta/n$. Also define a sequence $\{\bar{U}^n\}$ according to

$$\bar{U}^n(x) \doteq (J^n \log R)/n - (j \log R)/n \text{ for } x \in C_{J^n - j}^n \backslash C_{J^n - j - 1}^n, \quad j = 0, 1, \ldots, J^n.$$

Note that whenever $y_n \to y$, $\bar{U}^n(y_n) - \bar{U}^n(x_0) \to \bar{V}(y) \wedge \bar{V}(x_0) - \bar{V}(x_0)$. Note also that if $\bar{V}$ is a subsolution in the sense of Definition 16.12, then so is $\bar{V}(\cdot) \wedge \bar{V}(x_0)$. To simplify notation, we assume without loss that $\bar{V}(x) \leq \bar{V}(x_0)$ for all $x \in D$, and therefore for $x \in D$,

$$\left| (\bar{U}^n(x_0) - \bar{U}^n(x)) - (\bar{V}(x_0) - \bar{V}(x)) \right| \leq \frac{\log R}{n}. \qquad (16.27)$$

In particular, $\bar{U}^n(x_0) - \bar{U}^n(x) \to \bar{V}(x_0) - \bar{V}(x)$ for all $x \in D$. We now apply Theorems 16.6 and 16.7, with Δ replaced by Δ^n and $\bar{U}$ replaced by $n\bar{U}^n$, to the

Markov chain $\{X_i^n\}$. Following the same notation as in Chaps. 14 and 15, we denote the second moment of the estimator $E_{x_0}[(\gamma^n)^2]$ by $\mathfrak{S}^n(\bar{V})$. The corresponding initializing distribution λ_k^n is defined as $\lambda_k^n(l) = q_l^n / R^{J^n - k}$, with q_l^n defined by (16.5) but with J replaced by J^n. Theorems 16.6 and 16.7 then say that

$$
e^{-n\bar{U}^n(x_0)} E_{x_0} \left[\sum_{i=1}^{M^n} e^{n\bar{U}^n(X_{i-1}^n)} \left(E_{X_i^n} \left[1_B(X_{M^n}^n) \right] \right)^2 \right]
$$
$$
+ e^{-n\bar{U}^n(x_0)} E_{x_0} \left[e^{n\bar{U}^n(X_{M^n}^n)} 1_B(X_{M^n}^n) \right] \geq \mathfrak{S}^n(\bar{V}) \tag{16.28}
$$

and

$$
\mathfrak{S}^n(\bar{V}) \geq e^{-n\bar{U}^n(x_0)} E_{x_0} \left[e^{n\bar{U}^n(X_{M^n}^n)} 1_B(X_{M^n}^n) \right],
$$

where $\gamma^n = e^{-n\bar{U}^n(x_0)} \sum_{i=0}^{\infty} \int_{\mathbb{R}^d} 1_B(y) e^{-n\bar{U}^n(y)} \delta_{Z_i^n}(dy)$ is the estimator of (16.20) based on the importance function $\bar{U}^n$.

Theorem 16.15 below describes the asymptotic performance of the splitting scheme based on importance functions $\{\bar{U}^n\}$. As a consequence of this result, in Corollary 16.16, we will see that the decay rate

$$
\lim_{n\to\infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V})
$$

is bounded below by $V(x_0) + \bar{V}(x_0)$, where $V(x_0)$ is defined in (16.25), which is the same as the decay rate for an importance sampling scheme based on the same subsolution if it is sufficiently regular (see Theorem 15.10). We recall from (14.5) that when the large deviation limit holds, the decay rate of any unbiased splitting scheme is bounded above by $2V(x_0)$. In particular, if $\bar{V}(x_0) = V(x_0)$, we get the best possible decay rate $2V(x_0)$. Finally, in Theorem 16.18 below, we will show that

$$
\lim_{n\to\infty} \frac{1}{n} \log E_{x_0} \left[w^n \right] = 0.
$$

Thus the work associated with such a scheme grows subexponentially, and consequently, the decay rate of the work-normalized error is zero, which is the best possible rate.

It is easily checked that if $\bar{V}$ is not a subsolution, then at points where the subsolution property fails, the branching is supercritical, and hence in this case there exists $y \in D$ such that if $y_n \to y$, then

$$
\liminf_{n\to\infty} \frac{1}{n} \log E_{y_n} \left[w^n \right] > 0.
$$

It follows that importance functions that are not obtained from subsolutions should not be used to design schemes, since it is possible that the computational costs of such schemes will grow exponentially.

Recall that $x_0 \notin \bar{A} \cup B$, $\bar{V}$ is a subsolution as in Definition 16.12, and that as noted previously, we can assume $\bar{V}(x) \leq \bar{V}(x_0)$ for all $x \in D$. Recall also the definition of $K_{y,T}$ in (16.23).

Theorem 16.15 *Assume Conditions 16.9–16.11. Then for $x_0 \notin \bar{A} \cup B$,*

$$
\lim_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V})
$$
$$
= \inf_{\phi \in K_{x_0,T}, T < \infty} \left[\int_0^T L(\phi(r), \dot{\phi}(r)) dr + \bar{V}(\phi(0)) - \bar{V}(\phi(T)) \right]. \qquad (16.29)
$$

Proof We first consider the lower bound

$$
\liminf_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V})
$$
$$
\geq \inf_{\phi \in K_{x_0,T}, T < \infty} \left[\int_0^T L(\phi(r), \dot{\phi}(r)) dr + \bar{V}(\phi(0)) - \bar{V}(\phi(T)) \right], \qquad (16.30)
$$

which is based on (16.28). While there are two terms in (16.28), the second term can be treated in a similar manner as the first term, and so we focus on the first. This term is

$$
e^{-n\bar{U}^n(x_0)} E_{x_0} \left[\sum_{i=1}^{M^n} e^{n\bar{U}^n(X_{i-1}^n)} 1_B \left(X_{M^{1,i,n}}^{1,i,n} \right) 1_B \left(X_{M^{2,i,n}}^{2,i,n} \right) \right], \qquad (16.31)
$$

where $\{X_j^{k,i,n}\}_{j \geq i}$, $k = 1, 2$, are (conditionally) independent copies of $\{X_j^n\}_{j \geq i}$ that start at X_i^n at $j = i$, and $M^{k,i,n}$ are the corresponding escape times.

We claim that instead of the large deviation asymptotics of (16.31), it suffices to consider the large deviation asymptotics of

$$
e^{-n\bar{U}^n(x_0)} E_{x_0} \left[\sum_{i=1}^{\lfloor nT \rfloor} 1_{\{M^n \geq i\}} e^{n\bar{U}^n(X_{i-1}^n)} 1_B \left(X_{M^{1,i,n} \wedge \lfloor nT \rfloor}^{1,i,n} \right) 1_B \left(X_{M^{2,i,n} \wedge \lfloor nT \rfloor}^{2,i,n} \right) \right] \qquad (16.32)
$$

for some fixed and finite T. Assuming the claim, observe that there are no more than order-n terms in the expected value, and it suffices to obtain the desired bound on each of these terms. Let i_n index such a term. In obtaining a bound, we can assume without loss that i_n/n will converge to some limit $t \in [0, T]$, and to simplify notation, we write i for i_n. We first show that

$$
\liminf_{n \to \infty} -\frac{1}{n} \log e^{-n\bar{U}^n(x_0)} E_{x_0} \left[e^{n\bar{U}^n(X_{i-1}^n)} 1_{\{M^n \geq i\}} 1_B \left(X_{M^{1,i,n} \wedge \lfloor nT \rfloor}^{1,i,n} \right) 1_B \left(X_{M^{2,i,n} \wedge \lfloor nT \rfloor}^{2,i,n} \right) \right]
$$
$$
\geq \inf \left[\int_0^s L(\phi(r), \dot{\phi}(r)) dr + \bar{V}(x_0) - \bar{V}(\phi(s)) \right], \qquad (16.33)
$$

where the infimum is over all absolutely continuous ϕ such that $\phi(0) = x_0$ and $\phi(r) \notin A \cup B^\circ$ for $r \in (0, s)$ and $\phi(s) \in B$ for some $s \geq t$. Let $\hat{X}^n(t)$ be the continuous time trajectory that interpolates X_j^n up till i, and thereafter is a two-component process that interpolates $X_j^{1,i,n}$ and $X_j^{2,i,n}$ up until $\lfloor nT \rfloor$. It is straightforward, using the Markov property and the uniformity of the large deviation estimates with respect to initial conditions that is assumed in Condition 16.9, to check that $\{\hat{X}^n\}$ satisfies a large deviation property, and that the rate function (with obvious notation for a trajectory η that branches at time t into η^1 and η^2) is

$$\int_0^t L(\eta(r), \dot{\eta}(r))dr + \int_t^T L(\eta^1(r), \dot{\eta}^1(r))dr + \int_t^T L(\eta^2(r), \dot{\eta}^2(r))dr.$$

Since $\bar{V}$ is continuous and B is closed, we obtain the lower bound

$$\inf \left[\int_0^t L(\eta(r), \dot{\eta}(r))dr + \int_t^T L(\eta^1(r), \dot{\eta}^1(r))dr \right.$$
$$\left. + \int_t^T L(\eta^2(r), \dot{\eta}^2(r))dr + \bar{V}(x_0) - \bar{V}(\eta(t)) \right]$$

for the left side of (16.33), where the infimum is over all η such that $\eta(0) = x_0$ and $\eta(r) \notin A \cup B^\circ$ for $r \in (0, t]$ and $\eta^k, k = 1, 2$ such that $\eta^k(t) = \eta(t)$ and $\eta^k(r) \notin A \cup B^\circ$ for $r \in [t, s^k], s^k \in [t, T], \eta^k(s^k) \in B$. Without loss we can assume that the cost is zero after s^k and that $\eta^1 = \eta^2$ (which we relabel as η, and s^k as s). By the subsolution property,

$$\bar{V}(\eta(t)) \leq \int_t^s L(\eta(r), \dot{\eta}(r))dr + \bar{V}(\eta(s)),$$

which gives (16.33).

We now prove the claim. It remains to show that (16.31) has the same large deviation asymptotics as (16.32). To justify bounding the other random times by $\lfloor nT \rfloor$, we need to show that

$$\limsup_{T \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log e^{-n\bar{U}^n(x_0)} E_{x_0} \left[\sum_{i=1}^{M^n \wedge \lfloor nT \rfloor} e^{n\bar{U}^n(X_{i-1}^n)} \left(1_{\{\tau^{1,i,n} \geq nT\}} + 1_{\{\tau^{2,i,n} \geq nT\}} \right) \right]$$
$$= -\infty. \tag{16.34}$$

However, using (16.27), the expected value is bounded above by

$$2Re^{n2\|\bar{V}\|_\infty} \sum_{i=1}^{\lfloor nT \rfloor} P_{x_0} \left\{ \tau^{1,i,n} \geq nT \right\},$$

and thus (16.34), and therefore the claim, follows from Condition 16.10 [see (16.22)].

We now prove the upper bound

$$\limsup_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V})$$

$$\leq \inf_{\phi \in K_{x_0,T}, T < \infty} \left[\int_0^T L(\phi(r), \dot{\phi}(r)) dr + \bar{V}(\phi(0)) - \bar{V}(\phi(T)) \right]. \qquad (16.35)$$

Fix $\varepsilon \in (0, 1)$, let Γ denote the right-hand side of (16.35), and using Condition 16.11 choose an absolutely continuous ϕ, T and $\tau \in (0, T)$ such that $\phi(0) = x_0, \phi(\tau) \in B$, $\phi(t) \notin \bar{A} \cup B$ for all $t \in (0, \tau)$, $\phi(t) \in B^\circ$ for all $t \in (\tau, T]$, and

$$\int_0^T L(\phi(r), \dot{\phi}(r)) dr + \bar{V}(\phi(0)) - \bar{V}(\phi(\tau)) \leq \Gamma + \varepsilon. \qquad (16.36)$$

Let $\delta > 0$ satisfy $\|\phi(r) - x\| \geq \delta$ if $r \in [0, \tau - \delta)$ and $x \in \bar{A} \cup B$ and $\|\phi(r) - x\| \geq \delta$ if $r \in [\tau + \delta, T]$ and $x \notin B$. Let also $\Xi^n \doteq \{\sup_{0 \leq t \leq T} \|X^n(t) - \phi(t)\| < \delta\}$. Then using the large deviation lower bound for the third inequality and (16.36) for the fourth, we obtain

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathfrak{S}^n(\bar{V})$$

$$\geq \liminf_{n \to \infty} \frac{1}{n} \log e^{-n\bar{U}^n(x_0)} E_{x_0} \left[e^{n\bar{U}^n(X_{M^n}^n)} 1_B(X_{M^n}^n) \right]$$

$$\geq \liminf_{n \to \infty} \frac{1}{n} \log e^{-n\bar{U}^n(x_0)} E_{x_0} \left[e^{n\bar{U}^n(X_{M^n}^n)} 1_B(X_{M^n}^n) 1_{\Xi^n} \right]$$

$$\geq \inf_{y \in C_\delta} \bar{V}(y) - \bar{V}(x_0) + \liminf_{n \to \infty} \frac{1}{n} \log P_{x_0}\{\Xi^n\}$$

$$\geq \inf_{y \in C_\delta} \bar{V}(y) - \bar{V}(\phi(\tau)) - \Gamma - \varepsilon,$$

where $C_\delta \doteq \{y : \|y - \phi(t)\| < \delta$ for some $t \in [\tau - \delta, \tau + \delta]\}$. Since $\bar{V}$ and ϕ are continuous, we have $\inf_{y \in C_\delta} \bar{V}(y) - \bar{V}(\phi(\tau)) \to 0$ as $\delta \to 0$. Since $\varepsilon > 0$ and $\delta > 0$ are arbitrary, this proves the upper bound in (16.35). $\qquad \square$

Corollary 16.16 *Under same conditions and with the same notation as in Theorem 16.15,*

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V}) \geq V(x_0) + \bar{V}(x_0).$$

In particular, if $\bar{V}(x_0) = V(x_0)$, then

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V}) = 2V(x_0).$$

Proof Recall the set $K_{x_0,T}$ introduced in (16.23) and consider any $\phi \in K_{x_0,T}$ with $T \in (0, \infty)$. Since $\bar{V}$ is a subsolution, it follows that $\bar{V}(z) \leq 0$ for $z \in B$, and so

$$\int_0^T L(\phi(r), \dot{\phi}(r))dr + \bar{V}(\phi(0)) - \bar{V}(\phi(T)) \geq \int_0^T L(\phi(r), \dot{\phi}(r))dr + \bar{V}(x_0).$$

Taking the infimum over all $\phi \in K_{x_0,T}$ and $T \in (0, \infty)$, we have from (16.25) and (16.29) that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V}) \geq V(x_0) + \bar{V}(x_0),$$

proving the first statement in the corollary. On the other hand, as was argued previously,

$$\limsup_{n \to \infty} -\frac{1}{n} \log \mathfrak{S}^n(\bar{V}) \leq 2V(x_0).$$

The second statement in the corollary follows.   □

Remark 16.17 An examination of the proof shows that the greatest contribution to the second moment of the estimator is from the correlation of particles that make it to B and whose last common ancestor is located in one of the thresholds close to B.

Finally, we now show that the work associated with the splitting scheme based on the sequence $\{\bar{U}^n\}$ grows subexponentially.

Theorem 16.18 *Under same conditions and with the same notation as in Theorem 16.15,*

$$\lim_{n \to \infty} \frac{1}{n} \log E_{x_0}\left[w^n\right] = 0,$$

where w^n is defined by the right side of (16.11) replacing N_i by $N_i^n \doteq \int_D \bar{\delta}_{Z_i^n}(dx)$.

Proof We know from Theorem 16.5 that

$$E_{x_0}\left[w^n\right] = e^{n\bar{U}^n(x_0)} E_{x_0}\left[\sum_{i=0}^{M^n} e^{-n\bar{U}^n(X_i^n)}\right].$$

Exactly as in the proof of (16.30), it follows that the large deviation asymptotics of $E_{x_0}[w^n]$ are the same as those of

$$e^{n\bar{U}^n(x_0)} E_{x_0}\left[\sum_{i=0}^{M^n \wedge \lfloor nT \rfloor} e^{-n\bar{U}^n(X_i^n)}\right]$$

for some sufficiently large but finite T. The convergence $\bar{U}^n(y) - \bar{U}^n(x_0) \to \bar{V}(y) - \bar{V}(x_0)$ and the same line of argument as in the proof of (16.30) show that

$$\limsup_{n\to\infty} \frac{1}{n} \log e^{n\bar{U}^n(x_0)} E_{x_0} \left[\sum_{i=0}^{M^n \wedge \lfloor nT \rfloor} e^{-n\bar{U}^n(X_i^n)} \right]$$

$$\leq - \inf_{\phi \in K_{x_0,T}, T<\infty} \left[\int_0^T L(\phi(s), \dot{\phi}(s)) ds - (\bar{V}(x_0) - \bar{V}(\phi(T))) \right].$$

By the subsolution property (Definition 16.12), the quantity to be minimized is always nonnegative, and so the upper bound follows. Since $E_{x_0}[w^n] \geq 1$ for all n, the lower bound is automatic, which completes the proof. $\square$

Remark 16.19 Although the subsolution property implies a type of stability as asserted in Theorem 16.18, it could allow for polynomial growth of the number of particles. If in practice one observes that a large number of particles make it to B in the course of simulating a single sample, then one can consider the use of a *strict* subsolution, i.e., a function $\bar{V}$ that satisfies the boundary conditions and

$$\bar{V}(y) \leq \inf_{\phi \in K_{y,T}, T<\infty} \left[\int_0^T [L(\phi(s), \dot{\phi}(s)) - \varepsilon] ds + \bar{V}(\phi(T)) \right]$$

for some $\varepsilon > 0$. Because the value of $\bar{V}(x_0)$ is lowered slightly, there will be a slight increase in the second moment of the estimator. However, the strict inequality provides stronger control, and indeed, the expected number of particles and moments of the number of particles are bounded uniformly in n. See [76] for further discussion and examples. If phrased in terms of $\mathbb{H}$ as in Remark 16.14, the strict subsolution property means that $\mathbb{H}(x, D\bar{V}(x)) \geq \varepsilon$ for $x \in (A \cup B^\circ)^c$.

16.5   Splitting for Finite-Time Problems

By adding time as a state variable, finite-time problems such as those discussed in the context of importance sampling in Sect. 15.2 can also be be put into the RESTART framework. Thus the process $\{X_i^n\}$ is replaced by $\{(X_i^n, t_i^n)\}$, where

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x_0, \quad t_{i+1}^n = t_i^n + \frac{1}{n}, \quad t_0^n = 0.$$

Consider, for example, the estimation of $P_{x_0}\{X^n(T) \in B\}$. For this problem, it is assumed that the rare set $B \subset \mathbb{R}^d$ does not contain the terminal values $\phi(T)$ of zero-cost trajectories that start at x_0, and the typical behavior is $\{X^n(T) \in A\}$ with $A = B^c$. (Note that if we wish to continue the reduction to the time-independent case, then with the state space $\mathbb{R}^{d+1}$, we would call the rare set $B \times \{T\} \subset \mathbb{R}^{d+1}$ and the typical set $A \times \{T\}$.) The definition of subsolution becomes the following, with $\bar{K}_{y,t,T}$ the set of absolutely continuous trajectories ϕ with $\phi(t) = y$ and $\phi(T) \in B$.

Definition 16.20 A continuous function $\bar{V} : \mathbb{R}^d \times [0, T] \to \mathbb{R}$ is a subsolution if it is bounded from below and

$$\bar{V}(y, t) \leq \inf_{\phi \in \bar{K}_{y,t,T}} \left[\int_t^T L(\phi(s), \dot{\phi}(s)) ds + \bar{V}(\phi(T), T) \right]$$

for all $(y, t) \in \mathbb{R}^d \times [0, T)$, and $\bar{V}(z, T) \leq 0$ for $z \in B$.

One can add a "bounding" set as in Remark 15.4, which does not change the requirement for $\bar{V}$ to be a subsolution, except that $A \times \{T\}$ now also includes the points $D^c \times [0, T)$, and we restrict in the definition to $(y, t) \in D \times [0, T)$. As in Remark 16.13, a classical or piecewise classical subsolution in the sense of Definitions 14.1 and 14.2 is a subsolution in the sense of Definition 16.20.

A related problem of interest is to estimate the probability of escaping any time during the time interval, i.e., $P_{x_0}\{X^n(t) \in B$ for some $t \in [0, T]\}$. In this case, one should replace B in the time-independent setting by $B \times [0, T]$, and A by $A \times \{T\}$. The definition of subsolution is then the following.

Definition 16.21 A continuous function $\bar{V} : \mathbb{R}^d \times [0, T] \to \mathbb{R}$ is a subsolution if it is bounded from below and

$$\bar{V}(y, t) \leq \inf_{\phi \in \bar{K}_{y,t,s}, t \leq s \leq T} \left[\int_t^s L(\phi(s), \dot{\phi}(s)) ds + \bar{V}(\phi(s), s) \right]$$

for all $(y, t) \in \mathbb{R}^d \times [0, T)$ and $\bar{V}(z, t) \leq 0$ for $z \in B$ and $t \in [0, T]$.

As an elementary time-dependent example, we consider the case in which the $\{v_i(x)\}$ are $N(0, 1)$ (and thus independent of x), so that $H(\alpha) = \alpha^2/2$ and $L(\beta) = \beta^2/2$. With $B = [1, \infty)$ and $T = 1$,

$$\bar{V}(x, t) = -x + \frac{1}{2}t + \frac{1}{2}$$

is a subsolution with the optimal value at $(0, 0)$. Splitting thresholds as well as the start of a simulation with splitting rate $R = 3$ are depicted in Fig. 16.3.

16.5.1   Subsolutions for Analysis of Metastability

Suppose that x^* has the property that all zero-cost trajectories are attracted to x^* in the sense that for all $x_0 \in \mathbb{R}^d$, the properties $I_S(\phi) = 0$ for all S and $\phi(0) = x_0$ imply that $\phi(S) \to x^*$ as $S \to \infty$. Consider the issue of estimating $P_{x^*}\{X^n(T) \in B\}$. Assume for simplicity that B^c is bounded and define

$$W(x, y) = \inf [I_S(\phi) : \phi(0) = x, \phi(S) = y, S < \infty].$$

Fig. 16.3 Splitting
thresholds for
time-dependent problem
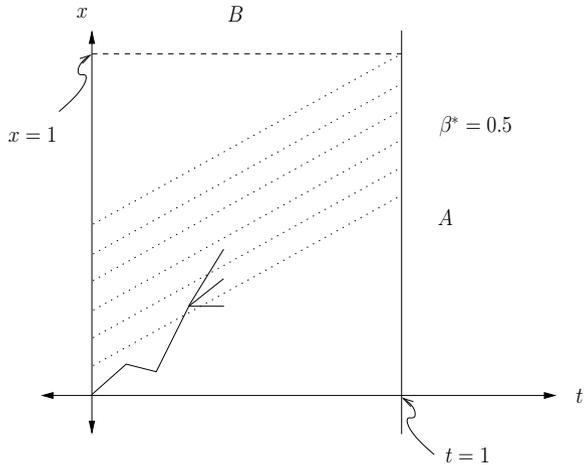
Then $W(x, y)$ is the Freidlin–Wentzell quasipotential [140] relative to the starting
point x. Suppose again for simplicity of presentation that $W(x^*, \cdot)$ is continuous. In
this context, a particularly convenient subsolution is that of the form

$$\bar{V}(y, t) = \bar{V}(y) = -W(x^*, y) + c.$$

Here $c \in \mathbb{R}$ is the largest value such that the boundary condition $-W(x^*, y) + c \le 0$
holds for all $y \in B$.

To see that $\bar{V}(y)$ is a subsolution, we note that $W(x, z)$ satisfies the dynamic
programming equation

$$W(x, z) = \inf_{y \in \mathbb{R}^d} [W(x, y) + W(y, z)],$$

from which it follows that for all $y \in \mathbb{R}^d$ and $c \in \mathbb{R}$,

$$-W(x^*, y) + c \le -W(x^*, z) + c + W(y, z).$$

The definition of $W(y, z)$ then gives [for all $\phi \in \bar{K}_{y,t,T}$ and with $z = \phi(T)$] that

$$\bar{V}(y) \le \bar{V}(\phi(T)) + \int_0^T L(\phi(s), \dot{\phi}(s))ds,$$

and therefore $\bar{V}$ is a subsolution. One can show that under appropriate conditions,
as $T \to \infty$,

$$\inf_{\phi \in \bar{K}_{x^*,t,T} : \phi(T) \in B} \left[\int_0^T L(\phi(s), \dot{\phi}(s))ds \right]$$

converges to $\bar{V}(x^*)$, and hence $\bar{V}(y)$ is a potentially useful subsolution for studying escape to B from a neighborhood of the attractor x^* at the end of a long time interval.

With regard to the problem of estimating $P_{x^*}\{X^n(t) \in B$ for some $t \in [0, T]\}$, $\bar{V}(y)$ again provides a (time-independent) subsolution with a nearly optimal value at x^* when T is large. The argument is similar to the case of $P_{x^*}\{X^n(T) \in B\}$ and is hence omitted.

16.6   Notes

Particle splitting methods originate with [166], and are further developed in [30]. A review of their application to rare-event problems appears in [145], as well as [223]. The RESTART algorithm, which is the focus of this chapter, was first presented in [241].

The main source for this chapter is [77], which uses a more general formulation and also phrases the assumptions to explicitly include queueing networks and expected values. Just as with importance sampling, some incorrect uses of large deviation asymptotics for the design of splitting schemes have been proposed, and a discussion on these issues can be found in [149].