A Large Deviation based Splitting Estimation of Power Flow Reliability

Wander S. Wadman, Daan T. Crommelin and Bert P. Zwart

November 23, 2015

Abstract

Given the continued integration of intermittent renewable generators in electrical power grids, connection overloads are of increasing concern for grid operators. The risk of an overload due to injection variability can be described mathematically as a barrier crossing probability of a function of a multidimensional stochastic process. Crude Monte Carlo is a well-known technique to estimate probabilities, but it may be computationally too intensive in this case as typical modern power grids rarely exhibit connection overloads. In this paper we derive an approximate rate function for the overload probability using results from large deviations theory. Based on this large deviations approximation, we apply a rare event simulation technique called splitting to estimate overload probabilities more efficiently than Crude Monte Carlo simulation.

We show on example power grids with up to eleven stochastic power injections that for a fixed accuracy Crude Monte Carlo would require tens to millions as many samples than the proposed splitting technique required. We investigate the balance between accuracy and workload of three splitting schemes, each based on a different approximation of the rate function. We justify the workload increase of large deviations based splitting compared to naive splitting — that is, splitting based on merely the Euclidean distance to the rare event set. For a fixed accuracy naive splitting requires over 60 times as much CPU time as large deviation based splitting, illustrating its computational advantage. In these examples naive splitting — unlike large deviations based splitting — requires even more CPU time than CMC simulation, illustrating its pitfall.

1 Introduction

Many modern societies have grown accustomed to a very reliable electricity supply by electrical power grids. However, substantial implementation of intermittent renewable generation, such as photovoltaic power or wind power, endangers grid reliability. Power imbalances caused by generation intermittency may cause grid stability constraints to be violated. Grid operators may even have to curtail power to avoid grid instability, whereby some demanded power consumption (or generation) is not delivered (or produced) in the end. As grid operators are obliged to keep reliability at a prescribed level, they must assess the probability of constraint violations.

A connection overload is an example of such a constraint violation. If an excessive amount of power flows through the connection – e.g., a transmission

line or cable —, the connections temperature will eventually exceed its allowed maximum. As a result the connection may get damaged or it may sag and loose tensile strength [Wan et al., 1999].

Various indices exist to assess grid reliability [Billinton and Li, 1994]. Many of these quantify the occurrence of constraint violations that lead to a power curtailment. Examples are the probability, expected duration or expected frequency of violations during a fixed time interval, or the expected induced energy curtailed due to these violations. This paper focuses on the probability that specific connections overload, given the distribution of the multidimensional stochastic process of all uncertain power injections. Overload probabilities are important indices for many long-term investment questions of grid operators. For example, the power grid may not meet reliability standards after a significant number of renewable generators are integrated in the power grid. Fast and accurate estimation of overload risks will enable otherwise computationally too intensive optimizations of power grid investments. Overload probability estimates will also improve short-term operational strategies as a grid operator can act on these statistics during the next day, hour or even minutes.

Monte Carlo simulation can be used to estimate connection overload probabilities. However, constraint violations causing power curtailments are rare events in modern power grids. In case of a time interval of one week, probabilities of 10^{-5} or even much smaller are not uncommon [Carden and Wintermantel, 2013, CEER, 2014]. Crude Monte Carlo (CMC) estimators for rare event probabilities may require a prohibitively large number of samples to achieve a fixed accuracy. Since one CMC sample already involves solving a large number of high dimensional nonlinear systems, CMC estimation is computationally too intensive for grid reliability analyses in general.

Rare event simulation techniques have been developed for accurate and efficient estimation of very small probabilities. Importance sampling and (multilevel) splitting are two well-known variants. In importance sampling, one samples from an alternative distribution, and the estimator is multiplied by a factor to correct for the induced bias [Rubino and Tuffin, 2009]. Crucial for variance reduction is to find a distribution that increases rare event occurrences. Splitting techniques do not change the distribution, but replicate sample paths whenever the rare event is presumed substantially more likely given the current chain state [Garvels, 2000, L'Ecuyer et al., 2006]. Crucial for variance reduction is a suitable importance function. An importance function assigns a value to the state of a sample path, and this value determines whether the path is replicated at the time of this state. Ideally, this function maps the system states to the probability of hitting the rare event set starting from that system state. This is very similar to importance sampling in the sense that information on typical paths to the rare event is desired.

In a simple case with a one-dimensional state space and an interval as the rare event set, the distance to the rare event set will serve as a suitable importance function [L'Ecuyer et al., 2006, Wadman et al., 2014]. In many other cases however, the choice for the importance function is more difficult. In particular, in a high dimensional state space – in this paper: multiple power injections modeled by stochastic processes — it is in general not immediately clear which typical path towards the rare event should be stimulated. [Glasserman et al., 1998] show the importance of choosing the levels in a way consistent with the most likely path to a rare set.

In this paper we derive the most likely path towards the rare event of a connection overload using results from large deviations theory. We model power injections as a vector of correlated Ornstein-Uhlenbeck processes in Section 2. Section 3 introduces the framework of a splitting simulation. We derive an expression for the most likely path towards a connection overload in Section 4. We show that the derivation of this path is exact up to a numerical optimization problem with one equality constraint. We show that this optimization problem becomes one-dimensional if we assume a linear mapping between power injections and the connection power flow. We use the corresponding decay rate to construct a suitable importance function for a splitting technique in Section 5. This importance function can be embedded in the framework of [Dean and Dupuis, 2009], of which the authors show asymptotically optimal performance. The performance of this importance function is also described in detail in [Miretskiy et al., 2012]. To reduce the workload we propose three approximations of the decay rate based on different numerical solvers of the optimization problem. We investigate the accuracy and workload of the three corresponding splitting schemes on different stochastic extensions of the IEEE-14 test case in Section 6. We compare the performance of the large deviations based splitting simulation to a naive splitting simulation based on merely the proximity of the rare event set in the injection space. We conclude in Section 7.

In the literature, rare event simulation techniques have been based on large deviations theory for many applications including finance, engineering, molecular biology and power systems [Guasoni and Robertson, 2008, Vanden-Eijnden and Weare, 2012, Dupuis et al., 2012, Nykvist, 2015]. [Nykvist, 2015, Chapter 5] designs an asymptotically optimal importance sampling scheme to estimate voltage collapse probabilities in a power grid by constructing subsolutions of the Hamilton-Jacobi equations associated with the large deviations of the system. Also splitting techniques have been applied to power systems. [Wang et al., 2011] and [Shortle, 2013] estimated small probabilities of instantaneous, cascading failures of grid components. Instead, in our work the rare event is a connection overloading during a certain time interval and the sources of uncertainty are the power injections. [Schlapfer and Mancarella, 2011] used splitting to estimate the probability of a transmission line temperature exceeding a critical value. Markov processes with a discrete state space are used whereas we use a continuous state space. Furthermore, our importance function is based on the asymptotic overload probability, and not on the proximity of the constraint state variable to its allowed maximum. In [Wadman et al., 2013] a splitting technique is applied to estimate various grid reliability indices. Although the current paper focuses on connection overload probabilities only, the large deviations approach enables accurate estimates even for high-dimensional state spaces (see Section 6.2) and in cases where the shortest path to the rare event is much less likely than the most likely path (see Section 6.4).

2 The power flow model

Let the following vector-valued stochastic process $\{\boldsymbol{X}^{\varepsilon}(t), t \geq 0\}$ denote n uncertain power injections of the power grid as function of time t, defined as a

multidimensional Ornstein-Uhlenbeck (OU) process:

$$d\boldsymbol{X}^{\varepsilon}(t) = \boldsymbol{D}(\boldsymbol{\mu} - \boldsymbol{X}^{\varepsilon}(t))dt + \sqrt{\varepsilon}\boldsymbol{L}d\boldsymbol{W}(t), \boldsymbol{X}^{\varepsilon}(0) = \boldsymbol{x}_0. \tag{1}$$

Here $\boldsymbol{D} := \operatorname{diag}(\theta_1, \ldots, \theta_n) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with mean-reverting terms $\theta_1, \ldots, \theta_n > 0$ on the diagonal. The vector of long-term means is denoted by $\boldsymbol{\mu} \in \mathbb{R}^n$, $\varepsilon > 0$ is a scalar, $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ is a lower triangular matrix with $\boldsymbol{\Sigma} := \boldsymbol{L}\boldsymbol{L}^{\top}$ the covariance matrix of $\boldsymbol{L}\boldsymbol{W}(1)$, and $\boldsymbol{W}(t)$ is a vector of i.i.d. standard Brownian motions. Then $X_i^{\varepsilon}(t)$ is clearly a one-dimensional OU process with mean-reverting term θ_i, long-term mean μ_i, volatility $\sqrt{\varepsilon\Sigma_{ii}}$ and initial value $x_{0,i}$. The injection pattern at node i will therefore deviate according to $\sqrt{\varepsilon\Sigma_{ii}}$ but will revert back to mean $\boldsymbol{\mu}_i$ with force θ_i.

We admit that the model of OU processes is partially chosen for its tractability. However, multiple generators and consumers are often connected to one node — particularly in transmission grids. In those cases assuming a Gaussian process for the sum of their injection patterns is reasonable and has been done before [Perninge et al., 2011, Leith et al., 2004]. Especially on shorter time scales, assuming OU processes can be justified since power injections are expected to deviate from and attract to some historical average. Furthermore, the model incorporates dependencies through $\boldsymbol{L}$ between different power injections, reflecting the correlation between the meteorological sources of renewable energy or between consumption at different nodes.

Appendix A shows that $\boldsymbol{X}^{\varepsilon}(t)$ follows a multivariate normal distribution. We should note that the state space of power injections is therefore unbounded in this model, whereas in fact this state space is bounded for existing generators and consumers. On a larger scale however, many sources of generation and consumption are aggregated per node, so using the central limit theorem their net power injection at time t can be modeled realistically by an unbounded (normal) distribution.

We define the function $p : \mathbb{R}^n \to \mathbb{R}$ that maps the power injections to the power flowing through a specific grid connection. A common choice for p involves the AC power flow equations, which are described in detail in [Grainger and Stevenson, 1994, Chapter 9]. In short, a nonlinear algebraic system of steady state equations relates the power injections at each grid node to the voltages at all nodes. To compute a connection power flow at some time t given the power injections at that time, this nonlinear algebraic system has to be solved numerically for the nodal voltages. Then Ohm's law and the definition of power will immediately yield the power flow through a connection.

Another choice for p is a linear function of the power injections to the power flow through the connection of interest:

$$p(\boldsymbol{x}) = \boldsymbol{v}^{\top}\boldsymbol{x}, \tag{2}$$

for some constant vector $\boldsymbol{v} \in \mathbb{R}^n$. The DC power flow equations form a well-known example [Seifi and Sepasian, 2011, Appendix A], but also for radial AC networks linear functions have been derived [Low, 2014]. Since the experiments in Section 6 assume DC power flow equations, we will briefly describe this linear model:

$$\boldsymbol{P}_{\text{connection}}(\boldsymbol{x}) = \bar{\boldsymbol{B}}\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{x}.$$

Here, $\boldsymbol{P}_{\text{connection}} \in \mathbb{R}^l$ is the vector of power flowing through all l grid connections and $\boldsymbol{x} \in \mathbb{R}^n$ is the vector of power injections at all n grid nodes. In diagonal matrix $\bar{\boldsymbol{B}} \in \mathbb{R}^{l \times l}$ each diagonal entry $\bar{B}_{kk}$ is the electrical susceptance of connection k. The bus-branch incidence matrix $\boldsymbol{A} \in \mathbb{R}^{l \times n}$ contains the grid topology since it is defined as:

$$A_{ij} = \begin{cases} 0 & \text{if branch } i \text{ is not connected to node } j, \\ 1 & \text{if branch } i \text{ is connected to node } j \text{ and power flow from node } j \text{ is positive}, \\ -1 & \text{if branch } i \text{ is connected to node } j \text{ and power flow to node } j \text{ is positive}. \end{cases}$$

Finally, the elements of susceptance matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ are defined as:

$$B_{ij} = \begin{cases} -X_{ij}^{-1} & \text{if } i \neq j, \\ \sum_{i \neq j}^{n} X_{ij}^{-1} & \text{if } i = j. \end{cases}$$

Here X_{ij} is the electrical reactance of connection (i, j), which is assumed to be infinite if this connection does not exist. We conclude that the power flow through connection k is indeed of the form (2), with $\boldsymbol{v}^\top$ equal to the k-th row of matrix $\bar{\boldsymbol{B}}\boldsymbol{A}\boldsymbol{B}^{-1}$ and $p(\boldsymbol{x})$ the k-th element of $\boldsymbol{P}_{\text{connection}}(\boldsymbol{x})$.

Most results in this paper are derived and experimentally tested assuming p to be linear. However, as some results also hold for nonlinear p, we assume for now only that p is a deterministic and continuous function of $\boldsymbol{x}$ that solves a system of steady state equations, and we will mention it in later sections whenever we further assume linearity as in (2).

We are interested in the overload probability

$$\gamma := \mathbb{P}\left\{ \sup_{\tau \in (0,T]} p(\boldsymbol{X}^\varepsilon(\tau)) \geq P_{\max} \right\}$$

before some time $T > 0$, where $P_{\max} > 0$ is the maximum allowed value of power flowing through the connection.[1] As $\boldsymbol{X}^\varepsilon$ and p are continuous, we have

$$\gamma = \mathbb{P}\left\{ \exists\ \tau \in (0, T] : p(\boldsymbol{X}^\varepsilon(\tau)) = P_{\max} \right\}. \tag{3}$$

We assume throughout this paper that no overload occurs if the power injections are equal to their expectation:

$$p(\mathbb{E}[\boldsymbol{X}^\varepsilon(t)]) < P_{\max} \text{ for all times } t. \tag{4}$$

Formula (31) in Appendix A shows that

$$\mathbb{E}[\boldsymbol{X}^\varepsilon(t)] = e^{-\boldsymbol{D}t}(\boldsymbol{x}_0 - \boldsymbol{\mu}) + \boldsymbol{\mu},$$

where the exponential is a matrix exponential. So basically, assumption (4) implies that there is neither an overload at the starting time, $p(\boldsymbol{x}_0) < P_{\max}$, nor under average circumstances, $p(\boldsymbol{\mu}) < P_{\max}$, nor under the most likely circumstances in between. This reflects the connection being well-dimensioned under normal circumstances. Therefore for vanishing ε, $\mathbb{P}\{p(\boldsymbol{X}^\varepsilon(t)) \geq P_{\max}\}$ goes to zero for all t. We conclude that for fixed T, γ goes to zero for vanishing ε, which is why γ is a rare event probability if ε is small.

[1] The probability $\mathbb{P}\left\{\inf_{\tau \leq T} p(\boldsymbol{X}^\varepsilon(\tau)) \leq -P_{\max}\right\}$ is equally important for the grid operator, but is omitted here as this is a problem of completely similar complexity.

3 The splitting technique

To estimate (3) using CMC simulation, we sample trajectories from the discretization $\boldsymbol{X}_t^\varepsilon$ of $\boldsymbol{X}^\varepsilon(t)$ in (1) and check if $p(\boldsymbol{X}_t^\varepsilon) > P_{\max}$ at some discrete time t. Then the CMC estimator

$$\hat{\gamma}_{\text{CMC}} := \frac{1}{N}\sum_{i=1}^{N} 1_{\{\exists\ \tau\in(0,T]: p(\boldsymbol{X}_\tau^\varepsilon)\geq P_{\max}\text{ in sample } i\}}$$

for γ is in principle biased since rare event occurrences between subsequent time steps are ignored. However, when a connection overloads it takes some lag time before the connection temperature reaches a value where damage or sagging occurs [Wadman et al., 2012]. Therefore, grid operators are typically interested in overloads of a certain minimum duration, and we assume throughout this paper that the discrete time steps are chosen sufficiently small to ignore the discretization bias.

CMC simulation is inefficient for simulating rare events. The squared relative error

$$\frac{\operatorname{Var}\hat{\gamma}_{\text{CMC}}}{\gamma^2} = \frac{\gamma(1-\gamma)}{\gamma^2 N} = \frac{1-\gamma}{\gamma N} \tag{5}$$

of the CMC estimator diverges to infinity as $\mathcal{O}(1/\gamma)$ when N is fixed and $\gamma \to 0$. Therefore, to estimate a very small probability using CMC simulation, one may need a prohibitively large number of samples. Multilevel splitting, or (importance) splitting, is a rare event simulation technique developed to decrease this computational burden. Details behind splitting can be found in for example [Garvels, 2000, Rubino and Tuffin, 2009, L'Ecuyer et al., 2006, Botev and Kroese, 2012]. To keep this paper self-contained we give a brief introduction, similar to that in [Wadman et al., 2014].

Any splitting technique starts by defining an importance function $h : [0,T]\times\mathbb{R}^n \mapsto \mathbb{R}$ that assigns a value to each chain state $(t,\boldsymbol{x})$. It is constructed such that $h(t,\boldsymbol{x}) \geq 1$ precisely when $(t,\boldsymbol{x})$ corresponds to a rare event occurrence and $h(0,\boldsymbol{x}_0) = 0$. Further, higher values of a suitable importance function correspond to a chain state from which the rare event is more likely. The interval $[0,1]$ is divided into m subintervals with intermediate thresholds $0 = l_0 < l_1 < \cdots < l_m = 1$. Let $T_k = \inf\{t > 0 : h(t,\boldsymbol{X}^\varepsilon(t)) \geq l_k\}$ be the first time of hitting the k-th level and $E_k = \{T_k < T\}$ the event that the k-th level is hit during $[0,T]$. Obviously, $\mathbb{P}(E_m)$ is the value of interest as it is equal to γ. Also, $\mathbb{P}(E_0) = 1$. Since $E_m \subset E_{m-1} \subset \cdots \subset E_0$, we can write

$$\gamma = \prod_{k=1}^{m} \mathbb{P}(E_k|E_{k-1}). \tag{6}$$

That is, γ is a product of m conditional probabilities $p_k := \mathbb{P}(E_k|E_{k-1})$, which we will estimate separately. Independent sample paths from the conditional distribution of the entrance state $(T_{k-1}, \boldsymbol{X}^\varepsilon(T_{k-1}))$ given E_{k-1} would give us an estimate for p_k. However, we do not know this distribution for levels $k > 1$, and we use its empirical distribution instead, obtained from samples of the previous level. We proceed recursively in this way, and at each level k we estimate p_k by

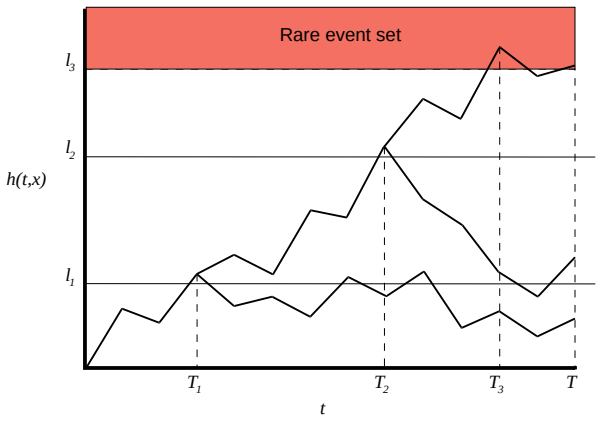

Figure 1: A minimal example of a splitting simulation.

the proportion $\hat{p}_k$ of sample paths for which E_k occurs (see Figure 1). Then the product

$$\hat{\gamma} := \prod_{k=1}^{m} \hat{p}_k \tag{7}$$

is an unbiased estimator for γ for several variants of the splitting technique, and Fixed Number of Successes (FNS) is one of them. FNS repeats generating sample paths at each level k until a prespecified number r_k hits of the next level are observed. The conditional probabilities p_k are estimated by $\hat{p}_k := (r_k - 1)/(n_k - 1)$, with n_k the number of samples generated at level k. [Amrein and Künsch, 2011] shows for this definition of $\hat{p}_k$ that (7) is indeed an unbiased estimator for γ. The authors also show that under ideal circumstances the squared relative error of the FNS estimator diverges as $\mathcal{O}((\log \gamma)^2)$ when $\gamma \to 0$. This squared logarithmic divergence rate is slower than the divergence rate of the CMC squared relative error in (5), illustrating the potential gain of splitting. We use the FNS splitting technique in the rest of this paper.

The choice for the importance function is crucial for variance reduction of the splitting estimator [Garvels, 2000, L'Ecuyer et al., 2006]. Intuitively, the importance function should 'reward good behavior' by splitting sample paths that are more likely to hit the rare event set. The levels should be chosen in a way consistent with the most likely path to the rare event set [Glasserman et al., 1998]. [Garvels et al., 2002] propose to use the importance function equal to (an increasing function of) the rare event probability given that one starts at the considered system state. As knowing this probability would defeat the point of using simulation, the lesson is to find an importance function that is close to this probability. We will use a result from large deviations theory to find an asymptotic probability of the rare event, in the limit of rarity parameter ε.

4 Results from large deviations theory

The following derivation is inspired by [Bosman et al., 2014]. Using the Freidlin–Wentzell theorem [Dembo and Zeitouni, 2009], the authors derive the decay

rate and most likely path of the limiting barrier crossing probability of a one-dimensional OU process before a fixed end time. We generalize this work to a function of multiple correlated OU processes (1), and use the result to construct a suitable importance function in Section 5.

Definition 1 *Probability measure $\mathbb{P}_\varepsilon$ on $(\mathcal{X}, \mathcal{B})$ is said to follow a large deviations principle (LDP) with a (lower semicontinuous) rate function $I : \mathcal{X} \to [0, \infty)$ if, for all $\Gamma \in \mathcal{B}$,*

$$-\inf_{x \in \Gamma^o} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon(\Gamma) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x),$$

with Γ^o and $\bar{\Gamma}$ the interior and closure, respectively, of Γ [Dembo and Zeitouni, 2009].

One may interpret this as $\mathbb{P}(\boldsymbol{X}^\varepsilon \in A) \sim \exp(-\inf\{I(x) : x \in A\}/\varepsilon)$ for small ε. Furthermore, rate function I is good if $\{x : I(x) \leq \alpha\}$ is a compact subset of $\mathcal{X}$ for all $\alpha \in [0, \infty)$. Then the Freidlin–Wentzell theorem asserts that $\{\boldsymbol{X}^\varepsilon\}$ satisfies a large deviation principle with good rate function $I_{0,\boldsymbol{x}_0}(\boldsymbol{x})$ defined by

$$I_{s,\boldsymbol{x}_s}(\boldsymbol{x}) = \begin{cases} \frac{1}{2}\int_s^T \mathcal{L}(\boldsymbol{x}, \boldsymbol{x}')dt & \text{if } \boldsymbol{x} \in H_{s,\boldsymbol{x}_s}, \\ \infty & \text{if } \boldsymbol{x} \notin H_{s,\boldsymbol{x}_s}, \end{cases} \quad \text{for } \; s \in [0,T), \boldsymbol{x}_s \in \mathbb{R}^n \tag{8}$$

with Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{u}^\top \boldsymbol{u}$, $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}') := \boldsymbol{L}^{-1}(\boldsymbol{x}' + (\boldsymbol{x} - \boldsymbol{\mu}))$, $\boldsymbol{x}' := \frac{d\boldsymbol{x}}{dt}$, and

$$H_{s,\boldsymbol{x}_s} = \left\{ \boldsymbol{x} : [s,T] \mapsto \mathbb{R}^n : x_i \in C[s,T], x_i(t) = x_{s,i} + \int_s^t \phi_i(z)dz, \phi_i \in L_2[s,T] \;\; \forall\, i \right\}.$$

We defined $I_{s,\boldsymbol{x}_s}, H_{s,\boldsymbol{x}_s}$ for the initial condition $\boldsymbol{x}(s) = \boldsymbol{x}_s$ for general $s \in [0,T)$ and $\boldsymbol{x}_s \in \mathbb{R}^n$ as this generalization of $s = 0$ and $\boldsymbol{x}_s = \boldsymbol{x}_0$ will become convenient in Section 5. In molecular dynamics — good rate function (8) is also known as the action functional [Vanden-Eijnden and Weare, 2012]. We define the minimum

$$I^*(s, \boldsymbol{x}_s) := \inf_{\tau \in (s,T], \boldsymbol{x} \in H_{s,\boldsymbol{x}_s} : p(\boldsymbol{x}(\tau)) = P_{\max}} I_{s,\boldsymbol{x}_s}(\boldsymbol{x})$$

of the good rate functions over all paths $\boldsymbol{x}$ that start in $\boldsymbol{x}_s$ at time $s \in [0,T)$ and that exhibit an overload $p(\boldsymbol{x}(\tau)) = P_{\max}$ at some time τ in the remaining time interval $(s,T]$. Then the contraction principle [Dembo and Zeitouni, 2009] implies

$$\lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{P}\left\{ \exists\, \tau \in (s,T] : p(\boldsymbol{X}^\varepsilon(\tau)) = P_{\max} \middle| \boldsymbol{X}^\varepsilon(s) = \boldsymbol{x}_s \right\} = I^*(s, \boldsymbol{x}_s).$$

Loosely speaking, the above asserts that the most likely path to the rare event set will become dominant as the rare event probability vanishes. More precisely, the decay rate of the vanishing probability converges to the good rate function of this most likely path. So in specific, the minimum good rate function of the probability in (3) converges to $I^*(0, \boldsymbol{x}_0)$, advocating the following approximation for small ε:

$$\gamma \approx e^{-I^*(0,\boldsymbol{x}_0)/\varepsilon}.$$

Therefore, the minimum good rate function is also called the decay rate. We should note that a subexponential factor times $\exp(-I^*(0,\boldsymbol{x}_0)/\varepsilon)$ would be a more accurate approximation, but as the exponential function dominates this factor as ε vanishes, we neglect it here [Touchette, 2009]. This approximation may serve as a first rough guess to distinguish connections with a significant overload probability. Furthermore, the most likely path will shed light on the typical combination of power injection paths that leads to an overload. We will use this approximation in Section 5 to construct a suitable importance function for a splitting simulation.

4.1 Minimizing the good rate function

We will write the decay rate $I^*(s,\boldsymbol{x}_s)$ as an minimization over $\tau \in (s,T]$ of an infimum $g_{s,\boldsymbol{x}_s}(\tau)$ for general τ:

$$I^*(s,\boldsymbol{x}_s) = \inf_{\tau\in(s,T]} g_{s,\boldsymbol{x}_s}(\tau), \qquad \text{with} \qquad g_{s,\boldsymbol{x}_s}(\tau) := \inf_{x\in H_{s,\boldsymbol{x}_s}: p(\boldsymbol{x}(\tau))=P_{\max}} I_{s,\boldsymbol{x}_s}(\boldsymbol{x}). \tag{9}$$

We use this formulation since we will first derive $g_{0,\boldsymbol{x}_0}(\tau)$ for general τ. Then we will show that the derivative $dg_{0,\boldsymbol{\mu}}/d\tau(\tau) < 0$ for all $\tau \in (s,T]$, implying that $g_{0,\boldsymbol{\mu}}(\tau)$ is smallest in $\tau = T$. This means that the most likely path from the mean to the rare event set enters the rare event set at the latest possible time.

Since the event $p(\boldsymbol{x}(\tau)) = P_{\max}$ only depends on the path $\boldsymbol{x}$ up until $t = \tau$, we have

$$\begin{aligned} g_{0,\boldsymbol{x}_0}(\tau) &= \inf_{x\in H_{0,\boldsymbol{x}_0}: p(\boldsymbol{x}(\tau))=P_{\max}} \frac{1}{2}\left(\int_0^\tau \mathcal{L}(\boldsymbol{x},\boldsymbol{x}')dt + \int_\tau^T \mathcal{L}(\boldsymbol{x},\boldsymbol{x}')dt\right) \\ &= \frac{1}{2}\inf_{x\in H_{0,\boldsymbol{x}_0}: p(\boldsymbol{x}(\tau))=P_{\max}} \int_0^\tau \mathcal{L}(\boldsymbol{x},\boldsymbol{x}')dt. \end{aligned} \tag{10}$$

Proposition 2 *The path $\boldsymbol{x}$ that minimizes (10) is of the form*

$$\boldsymbol{x}(t) = (\boldsymbol{V}e^{\boldsymbol{D}t} - e^{-\boldsymbol{D}t}\boldsymbol{V})\boldsymbol{c} + e^{-\boldsymbol{D}t}(\boldsymbol{x}_0 - \boldsymbol{\mu}) + \boldsymbol{\mu}, \tag{11}$$

with $\boldsymbol{c} \in \mathbb{R}^n$ such that $p(\boldsymbol{x}(\tau)) = P_{\max}$, and matrix $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ given by

$$V_{ij} = \frac{\Sigma_{ij}}{\theta_i + \theta_j}. \tag{12}$$

Proof 3 *See Appendix B.*

Proposition 2 specifies the most likely path $\boldsymbol{x}$ up to a constant vector $\boldsymbol{c} \in \mathbb{R}^n$ that fulfils $p(\boldsymbol{x}(\tau)) = P_{\max}$. Since $p(\boldsymbol{x}(\tau)) = P_{\max}$ is only one equation whereas the degrees of freedom for $\boldsymbol{c}$ is n, the Euler-Lagrange equations are in general not sufficient to find infimum (10). Therefore, we substitute (11) in (10) and further minimize the resulting objective function under the constraint $p(\boldsymbol{x}(\tau)) = P_{\max}$.

That is, since $\boldsymbol{x}'(t) = (\boldsymbol{V}\boldsymbol{D}e^{\boldsymbol{D}t} + \boldsymbol{D}e^{-\boldsymbol{D}t}\boldsymbol{V})\boldsymbol{c} - \boldsymbol{D}e^{-\boldsymbol{D}t}(\boldsymbol{x}_0 - \boldsymbol{\mu})$, we have

$$\begin{aligned}\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}') &= \boldsymbol{L}^{-1}\left(\left(\boldsymbol{V}\boldsymbol{D}e^{-\boldsymbol{D}t} + \boldsymbol{D}e^{-\boldsymbol{D}t}\boldsymbol{V} + \boldsymbol{D}\boldsymbol{V}e^{\boldsymbol{D}t} - \boldsymbol{D}e^{\boldsymbol{D}t}\boldsymbol{V}\right)\boldsymbol{c} + \boldsymbol{D}\mu - \boldsymbol{D}\mu\right) \\ &= \boldsymbol{L}^{-1}\left(\boldsymbol{V}\boldsymbol{D} + \boldsymbol{D}\boldsymbol{V}\right)e^{\boldsymbol{D}t}\boldsymbol{c} \\ &= \boldsymbol{L}^{-1}\boldsymbol{\Sigma}e^{\boldsymbol{D}t}\boldsymbol{c} \\ &= \boldsymbol{L}^{\top}e^{\boldsymbol{D}t}\boldsymbol{c}.\end{aligned}$$

In the third equality, $\boldsymbol{V}\boldsymbol{D} + \boldsymbol{D}\boldsymbol{V} = \boldsymbol{\Sigma}$ follows directly from the definitions of $\boldsymbol{V}$ and $\boldsymbol{D}$. The objective function in (10) becomes

$$\begin{aligned}\int_0^\tau \mathcal{L}(\boldsymbol{x}, \boldsymbol{x}')dt = \int_0^\tau \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}')^\top \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}')dt &= \int_0^\tau \boldsymbol{c}^\top e^{\boldsymbol{D}t}\boldsymbol{L}\boldsymbol{L}^\top e^{\boldsymbol{D}t}\boldsymbol{c}\, dt \\ &= \boldsymbol{c}^\top \int_0^\tau e^{\boldsymbol{D}t}\boldsymbol{\Sigma}e^{\boldsymbol{D}t}dt\ \boldsymbol{c} \\ &= \boldsymbol{c}^\top\left(e^{\boldsymbol{D}\tau}\boldsymbol{V}e^{\boldsymbol{D}\tau} - \boldsymbol{V}\right)\boldsymbol{c} \\ &= \boldsymbol{c}^\top e^{\boldsymbol{D}\tau}\left(\boldsymbol{V} - e^{-\boldsymbol{D}\tau}\boldsymbol{V}e^{-\boldsymbol{D}\tau}\right)e^{\boldsymbol{D}\tau}\boldsymbol{c} \\ &= \boldsymbol{c}^\top e^{\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))e^{\boldsymbol{D}\tau}\boldsymbol{c}.\end{aligned}$$

We used symmetry of the diagonal matrix exponential in the second equality and the integral in the fourth equality is easily derived elementwise. The covariance identity in the last equality, where $\boldsymbol{X}^1$ denotes X^ε for $\varepsilon = 1$, is shown elementwise in (32) in Appendix A. After substitution in (10) we obtain the minimization problem

$$g_{0,\boldsymbol{x}_0}(\tau) = \inf_{\boldsymbol{c}\in\mathbb{R}^n\ :\ p(\boldsymbol{x}_{\boldsymbol{c}}(\tau))=P_{\max}} \frac{1}{2}\boldsymbol{c}^\top e^{\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))e^{\boldsymbol{D}\tau}\boldsymbol{c} \tag{13}$$

over $\boldsymbol{c}$, where we introduced $\boldsymbol{x}_{\boldsymbol{c}}(t) := \boldsymbol{x}(t)$ as defined in (11) to emphasize the dependence on $\boldsymbol{c}$. Note that the objective function is quadratic in $\boldsymbol{c}$: if the constraint is linear, this optimization problem becomes a quadratic programming problem.

4.2 Starting from the mean the rare event is most likely at the end time

Now assume $\boldsymbol{x}_0 = \boldsymbol{\mu}$ — i.e., all processes start at their long-term mean — and define

$$\boldsymbol{b} := \left(\boldsymbol{V}e^{\boldsymbol{D}\tau} - e^{-\boldsymbol{D}\tau}\boldsymbol{V}\right)\boldsymbol{c} = \operatorname{Cov}(\boldsymbol{X}^1(\tau))e^{\boldsymbol{D}\tau}\boldsymbol{c}.$$

In this case minimization problem (13) becomes

$$g_{0,\boldsymbol{\mu}}(\tau) = \frac{1}{2}\inf_{\boldsymbol{b}\in\mathbb{R}^n : p(\boldsymbol{b}+\boldsymbol{\mu})=P_{\max}} \boldsymbol{b}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\boldsymbol{b}.$$

The optimal value $\boldsymbol{b}_{\text{opt}}$ for $\boldsymbol{b}$ clearly solves $p(\boldsymbol{b}_{\text{opt}} + \boldsymbol{\mu}) = P_{\max}$ so it does not depend on τ. The corresponding decay rate becomes

$$I^*(0, \boldsymbol{\mu}) = \frac{1}{2}\inf_{\tau\in(0,T]} \boldsymbol{b}_{\text{opt}}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\boldsymbol{b}_{\text{opt}}. \tag{14}$$

We differentiate the objective function with respect to τ:

$$\begin{aligned}\frac{d}{d\tau}\boldsymbol{b}_{\text{opt}}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\boldsymbol{b}_{\text{opt}} &= -\boldsymbol{b}_{\text{opt}}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\frac{d\operatorname{Cov}(\boldsymbol{X}^1(\tau))}{d\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\boldsymbol{b}_{\text{opt}} \\ &= -\boldsymbol{b}_{\text{opt}}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}e^{-\boldsymbol{D}\tau}\boldsymbol{\Sigma}e^{-\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}\boldsymbol{b}_{\text{opt}}.\end{aligned} \tag{15}$$

The first equality uses the identity for the derivative of a matrix inverse. The second equality holds since $\operatorname{Cov}(\boldsymbol{X}^1(\tau)) = \boldsymbol{V} - e^{-\boldsymbol{D}\tau}\boldsymbol{V}e^{-\boldsymbol{D}\tau}$ (see (32)) so its derivative is $e^{-\boldsymbol{D}\tau}(\boldsymbol{D}\boldsymbol{V} + \boldsymbol{V}\boldsymbol{D})e^{-\boldsymbol{D}\tau} = e^{-\boldsymbol{D}\tau}\Sigma e^{-\boldsymbol{D}\tau}$. We continue using the property that for an invertible matrix A and positive definite matrix B, $A^\top BA$ is positive definite too. Since $\boldsymbol{\Sigma}$ is a covariance matrix it is positive semi-definite, and without loss of generality we can assume it is positive definite. As $e^{-\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}$ is clearly invertible, the matrix

$$\operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}e^{-\boldsymbol{D}\tau}\boldsymbol{\Sigma}e^{-\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))^{-1}$$

is positive definite, so (15) is strictly negative for general $\boldsymbol{b}_{\text{opt}}$. Hence, the objective function in (14) strictly decreases in τ, so $\tau^* = T$ minimizes it:

$$I^*(0, \boldsymbol{\mu}) = g_{0,\boldsymbol{\mu}}(T).$$

We conclude that when the process start from the mean the most likely time to hit the rare event set is the last possible time T.

4.3 Quadratic programming assuming linear power flow equations

For the remainder of this paper we assume constraint function p to be linear, implying linear power flow equations. Then the minimization problem (13) has a closed-form solution: assuming $p(\boldsymbol{x}) = \boldsymbol{v}^\top\boldsymbol{x}$ as in (2), optimization program (13) becomes

$$g_{0,\boldsymbol{x}_0}(\tau) = \inf_{\boldsymbol{c}\in\mathbb{R}^n \,:\, \boldsymbol{v}^\top \operatorname{Cov}(\boldsymbol{X}^1(\tau))e^{\boldsymbol{D}\tau}\boldsymbol{c}=a} \frac{1}{2}\boldsymbol{c}^\top e^{\boldsymbol{D}\tau}\operatorname{Cov}(\boldsymbol{X}^1(\tau))e^{\boldsymbol{D}\tau}\boldsymbol{c},$$

with $a = a(\tau) := P_{\max} - \boldsymbol{v}^\top\left(\boldsymbol{\mu} + e^{-\boldsymbol{D}\tau}(\boldsymbol{x}_0 - \boldsymbol{\mu})\right)$. The minimizer $\boldsymbol{c}^*$ of this convex quadratic programming problem with one linear constraint is the solution of

$$\begin{pmatrix} e^{\boldsymbol{D}\tau}\operatorname{Cov}\boldsymbol{X}^1(\tau)e^{\boldsymbol{D}\tau} & e^{\boldsymbol{D}\tau}\operatorname{Cov}\boldsymbol{X}^1(\tau)\boldsymbol{v} \\ \boldsymbol{v}^\top\operatorname{Cov}\boldsymbol{X}^1(\tau)e^{\boldsymbol{D}\tau} & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{c}^* \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ a \end{pmatrix},$$

with $\lambda \in \mathbb{R}$ a (redundant) Lagrange multiplier [Murty, 2009]. Using an identity for the inverse of a block matrix, we obtain the minimizer

$$\boldsymbol{c}^* = a\frac{e^{-\boldsymbol{D}\tau}\boldsymbol{v}}{\boldsymbol{v}^\top\operatorname{Cov}\boldsymbol{X}^1(\tau)\boldsymbol{v}} \tag{16}$$

with corresponding minimum

$$g_{0,\boldsymbol{x}_0}(\tau) = \frac{1}{2}\frac{a^2}{\boldsymbol{v}^\top\operatorname{Cov}\boldsymbol{X}^1(\tau)\boldsymbol{v}}. \tag{17}$$

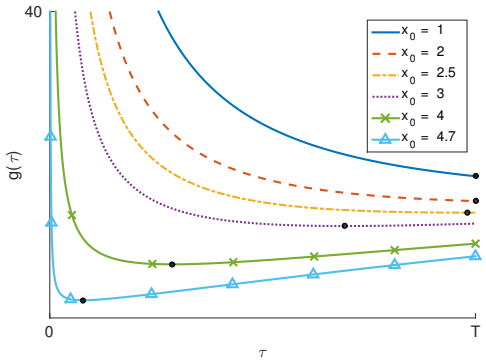

Figure 2: For values of $\boldsymbol{x}_0$ relatively close to rare event border $P_{\max}/\boldsymbol{v} = 5$ compared to $\boldsymbol{\mu} = 1$, the hitting time τ that minimizes decay rate $g(\tau)$ given a hit at τ is not the end time T. Black dots indicate minima $\tau^* = \arg\inf g(\tau)$ over the interval $[0, T]$.

By differentiating $g_{0,\boldsymbol{x}_0}$ to τ,

$$\frac{dg_{0,\boldsymbol{x}_0}}{d\tau} = \frac{1}{2}\frac{2a\boldsymbol{v}^\top \operatorname{Cov} \boldsymbol{X}^1(\tau)\boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{D}e^{-\boldsymbol{D}\tau}(\boldsymbol{x}_0 - \boldsymbol{\mu}) - a^2\boldsymbol{v}^\top e^{-\boldsymbol{D}\tau}\boldsymbol{\Sigma}e^{-\boldsymbol{D}\tau}\boldsymbol{v}}{(\boldsymbol{v}^\top \operatorname{Cov} \boldsymbol{X}^1(\tau)\boldsymbol{v})^2},$$

we confirm the result of Section 4.2 that if $\boldsymbol{x}_0 = \boldsymbol{\mu}$ the end time is the most likely time for the rare event to occur: again using properties of positive definite matrices, it is readily checked that $g_{0,\boldsymbol{\mu}}(\tau)$ indeed decreases in τ, so the decay rate becomes

$$I^*(0, \boldsymbol{\mu}) = g_{0,\boldsymbol{\mu}}(T) = \frac{1}{2}\frac{(P_{\max} - \boldsymbol{v}^\top \boldsymbol{\mu})^2}{\boldsymbol{v}^\top \operatorname{Cov} \boldsymbol{X}^1(T)\boldsymbol{v}}. \tag{18}$$

For general $\boldsymbol{x}_0 \in \mathbb{R}^n$ however, we do not have this guarantee. Even in the one-dimensional case $n = 1$, one can easily derive that the minimum of (17) is attained at

$$\tau^* = -\boldsymbol{D}^{-1} \log \frac{\boldsymbol{x}_0 - \boldsymbol{\mu}}{P_{\max}/\boldsymbol{v} - \boldsymbol{\mu}}. \tag{19}$$

This root is not defined if $\boldsymbol{x}_0 = \boldsymbol{\mu}$, but for any fixed T, sufficiently close values of $\boldsymbol{x}_0$ to $P_{\max}/\boldsymbol{v}$ this minimizer τ^* will be smaller than T. Figure 2 illustrates such counterexamples of $\boldsymbol{x}_0$ for $n = \boldsymbol{D} = \boldsymbol{\mu} = \boldsymbol{\Sigma} = \boldsymbol{v} = T = 1$ and $P_{\max} = 5$: indeed $\tau^* < T$ for $\boldsymbol{x}_0 \in [2.4715, 5]$, where the critical value 2.4715 is derived by solving (19) for $\boldsymbol{x}_0$ assuming $\tau^* = T$ and all other parameters as given.

5 A large deviations based importance function

In this section we generalize the results of Section 4 by conditioning on $\boldsymbol{X}^\varepsilon(s) = \boldsymbol{x}_s$ for general $s \in (0, T]$ instead of on $\boldsymbol{X}^\varepsilon(0) = \boldsymbol{x}_0$. We are interested in the decay rate $I^*(s, \boldsymbol{x}_s)$ of the limiting probability that, given $\boldsymbol{X}^\varepsilon(s) = \boldsymbol{x}_s$ at time s, the rare event $p(\boldsymbol{X}^\varepsilon(\tau)) \geq P_{\max}$ will occur at some time $\tau \in (s, T]$ in

the remaining time domain. We will use $I^*(s, \boldsymbol{x}_s)$ to compute an approximate probability to hit the rare event given a realized chain state. In turn, we will use this proxy as importance function in a splitting technique — i.e., to decide whether or not to split the sample path at the corresponding time step.

One can easily derive — i.e., completely analogous to (9)-(11) —, that the most likely path from $\boldsymbol{X}^\varepsilon(s) = \boldsymbol{x}_s$ to the rare event is of the form

$$\boldsymbol{x}(t) = (\boldsymbol{V}e^{\boldsymbol{D}t} - e^{-\boldsymbol{D}(t-s)}\boldsymbol{V}e^{\boldsymbol{D}s})\boldsymbol{c} + e^{-\boldsymbol{D}(t-s)}(\boldsymbol{x}_s - \boldsymbol{\mu}) + \boldsymbol{\mu} \tag{20}$$

for $t \in [s, T]$. Likewise, for general $s \in (0, T]$ the decay rate

$$I^*(s, \boldsymbol{x}_s) = \inf_{\tau \in (s,T]} g_{s,\boldsymbol{x}_s}(\tau), \tag{21}$$

with

$$g_{s,\boldsymbol{x}_s}(\tau) = \frac{1}{2} \inf_{\boldsymbol{c} \in \mathbb{R}^n : p(\boldsymbol{x}(\tau)) = P_{\max}} \boldsymbol{c}^\top e^{\boldsymbol{D}\tau} \operatorname{Cov}(\boldsymbol{X}^1(\tau - s)) e^{\boldsymbol{D}\tau} \boldsymbol{c} \tag{22}$$

can be derived analogously to (9)-(13). If we would assume linear power flow equations $p(\boldsymbol{x}) = \boldsymbol{v}^\top \boldsymbol{x}$ as in Section 4.3, the latter minimization has closed-form solution

$$g_{s,\boldsymbol{x}_s}(\tau) = \frac{1}{2} \frac{\left(P_{\max} - \boldsymbol{v}^\top \boldsymbol{\mu} - \boldsymbol{v}^\top e^{-\boldsymbol{D}(\tau - s)}(\boldsymbol{x}_s - \boldsymbol{\mu})\right)^2}{\boldsymbol{v}^\top \operatorname{Cov} \boldsymbol{X}^1(\tau - s) \boldsymbol{v}}. \tag{23}$$

To compute rare event probability γ using splitting, we should write it as in (6), a product of conditional probabilities. We will show now how to define these conditional probabilities in terms of the decay rate (21). First note that we can write the rare event probability of interest

$$\gamma = \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) = 0\right\}$$

as the probability that one arrives at a system state $(s, \boldsymbol{X}^\varepsilon(s))$ from where it takes 'zero Brownian effort' to arrive in the rare event set. Second, we can write the above in the form $\mathbb{P}(A|B)P(B)$, that is, as

$$\begin{aligned}\gamma = \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) = 0 \;\middle|\; \exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) < \alpha I^*(0, \boldsymbol{x}_0)\right\} \\ \times \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) < \alpha I^*(0, \boldsymbol{x}_0)\right\},\end{aligned}$$

for any threshold $\alpha \in (0, 1)$. This equality obviously holds since the condition $\{\exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) < \alpha I^*(0, \boldsymbol{x}_0)\}$ is a subset of the rare event $\{\exists s \in (0, T] : I^*(s, \boldsymbol{X}^\varepsilon(s)) = 0\}$. One may interpret $I^*(s, \boldsymbol{X}^\varepsilon(s))$ as a measure for the 'Brownian effort' required to arrive at the rare event set; so for $\alpha = 1/2$, the condition corresponds to the process being 'halfway towards the rare event set'.

We iterate this decomposition by first defining thresholds $0 =: l_0 < l_1 < \cdots < l_m := 1$ and events

$$E_k := \left\{\exists s \in (0, T] : 1 - \frac{I^*(s, \boldsymbol{X}^\varepsilon(s))}{I^*(0, \boldsymbol{x}_0)} \geq l_k\right\},$$

for $k = 0, \ldots, m$. Then since $\mathbb{P}\{E_0\} = 1$, $\mathbb{P}\{E_m\} = \gamma$ and $E_0 \supset \ldots \supset E_m$, the decomposition in (6) holds. This decomposition naturally suggests the large deviations based importance function $h : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$h_{\mathrm{LD}}(t, \boldsymbol{x}) := 1 - \frac{I^*(t, \boldsymbol{x})}{I^*(0, \boldsymbol{x}_0)}. \tag{24}$$

In [Dean and Dupuis, 2009], sufficient conditions are derived for an asymptotically optimal performance of a given importance function. They consider a probability of hitting a rare event set B before entering another set A. We can fit this setting with $B := \{p(y_{1,\dots,n}) \geq P_{\max}\} \cup \{t \leq T\}$, $A := \{t > T\}$ and stochastic process $\boldsymbol{y} := (\boldsymbol{X^\varepsilon(t)}, t)$. The authors show that under appropriate conditions, the asymptotic decay rate of the second moment of the splitting estimator $\hat{\gamma}^\varepsilon$ is optimally

$$\lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{E}[(\hat{\gamma}^\varepsilon)^2] = 2I^*(\boldsymbol{y}_0). \tag{25}$$

One condition for this optimality is that the function $\bar{W}(\boldsymbol{y}) := \mathbb{E}[r](1-h(\boldsymbol{y}))/\Delta$, with r the number of splitting particles, Δ the level size, satisfies

$$\bar{W}(\boldsymbol{y}) \leq 0 \qquad \text{for all } \boldsymbol{y} \in B, \tag{26}$$

$$\bar{W}(\boldsymbol{y_1}) - \bar{W}(\boldsymbol{y_2}) \leq \inf_{\boldsymbol{f},t:\boldsymbol{f}(0)=\boldsymbol{y_1},\boldsymbol{f}(t)=\boldsymbol{y_2}} \int_0^t \mathcal{L}(\boldsymbol{f}, \boldsymbol{f}')du. \quad \text{for all } \boldsymbol{y_1}, \boldsymbol{y_2} \notin A \cup B. \tag{27}$$

[Dean and Dupuis, 2009] call such a function a subsolution as it is the subsolution of the related Hamilton-Jacobi-Bellman equations. For $\bar{W}$ to satisfy both inequalities one requires an importance function h of the form $1-h(\boldsymbol{y}) \propto I^*(\boldsymbol{y})$. In this case the first inequality holds as $h(\boldsymbol{y}) > 1$ for all rare event set elements $\boldsymbol{y} \in B$. The second inequality can be written as a triangle inequality: the minimum good rate function of a path from point $\boldsymbol{y}$ to B is not larger when going directly than when traversing via point $\boldsymbol{y}$. Along these lines [Miretskiy et al., 2012] choose an importance function equal to the exponential decay rate to estimate a probability of first entrance into a rare set, and they prove asymptotic efficiency of their proposed Fixed Splitting scheme. The importance function (24) is similar to that in [Miretskiy et al., 2012].

5.1 Approximation of the decay rate: 3 algorithms

Assuming linear power flow equations $p(\boldsymbol{x}) = \boldsymbol{v}^\top \boldsymbol{x}$, computing $I^*(t, \boldsymbol{X}^\varepsilon(t))$ requires finding the optimal τ in (21)-(22). Although the search space of the optimization is one-dimensional, the optimization is required at every time step of every sample path in the splitting simulation. The associated workload will therefore challenge the computational advantage of rare event simulation as compared to CMC. To reduce the computational burden we define importance function

$$h_{\mathrm{LD},i} := 1 - \frac{I_i^*(t, \boldsymbol{x})}{I_i^*(0, \boldsymbol{x}_0)}$$

using the following three approximations $I_1^*(t, \boldsymbol{x}), I_2^*(t, \boldsymbol{x}), I_3^*(t, \boldsymbol{x})$ of decay rate $I^*(t, \boldsymbol{x})$:

Approximation 4 *We assume that the most likely time τ^* to enter the rare event set given the current state $(s, \boldsymbol{X}^\varepsilon(s))$ is one of the discrete time steps $s + \Delta, s + 2\Delta \dots, T$ of the discretization of (1). I_1^* denotes the corresponding decay rate approximation and $\hat{\gamma}_1$ the corresponding splitting estimator.*

The assumption in this approximation is reasonable for small step size Δ, and practical for any discrete time implementation of the involved stochastic processes. It reduces the optimization to computing $g_{s,\boldsymbol{x}_s}(\tau)$ for all $\tau = s+\Delta, s+2\Delta, \ldots, T$. As this assumption is weaker than those in the two subsequent approximations, the relative error of the accompanying splitting estimator will serve as a benchmark and thus it will be compared to that of a CMC estimator in Section 6.

Approximation 5 *For any $s < T$, we assume that the most likely time τ^* to enter the rare event set given the current state $(s, \boldsymbol{X}^\varepsilon(s))$ is end time T. I_2^* denotes the corresponding decay rate approximation and $\hat{\gamma}_2$ the corresponding splitting estimator.*

The assumption $\tau^* = T$ avoids the optimization problem (21)-(22) at each time step s. In Section 4.2 we have proven that this assumption is true if $\boldsymbol{x}_s = \boldsymbol{\mu}$. We expect that this assumption is reasonable for states $\boldsymbol{x}_s$ 'relatively close to $\boldsymbol{\mu}$' (see Figure 2). However, for sample paths that approach the rare event set relatively soon in the simulation, the end time T may be a suboptimal hitting time as then the mean-reverting force of the OU processes will require a very unlikely Brownian motion for a relatively long time.

Approximation 6 *We assume that the most likely state $(\tau^*, \boldsymbol{X}^\varepsilon(\tau^*))$ to enter the rare event set given the current state $(s, \boldsymbol{X}^\varepsilon(s))$ is independent of $(s, \boldsymbol{X}^\varepsilon(s))$. So for example if $\boldsymbol{x}_0 = \boldsymbol{\mu}$, for all $(s, \boldsymbol{X}^\varepsilon(s))$ the optimal rare event entrance state is $(T, \boldsymbol{x}_T)$ for some constant $\boldsymbol{x}_T \in \mathbb{R}^n$. I_3^* denotes the corresponding decay rate approximation and $\hat{\gamma}_3$ the corresponding splitting estimator.*

This assumption will hold exactly if a sample path is equal to the most likely path, since then the most likely rare event entrance state $(T, \boldsymbol{x}_T)$ given current state $(s, \boldsymbol{X}^\varepsilon(s))$ will be constant in time s. Although a sample path will in general not be exactly equal to the most likely path, the splitting technique will stimulate those paths that are close, so we may expect the most likely rare event entrance state to be close to $(T, \boldsymbol{x}_T)$, justifying the assumption.

We now derive the optimal hitting point $\boldsymbol{x}_T$ assuming $\boldsymbol{x}_0 = \boldsymbol{\mu}$. As shown in Section 4.2, $\tau^* = T$ is the most likely hitting time at $(0, \boldsymbol{X}^\varepsilon(0))$, explaining the notation $\boldsymbol{x}_T$ instead of $\boldsymbol{x}_{\tau^*}$. Expressions (11) and (16) then yield the most likely chain state

$$\begin{aligned}\boldsymbol{x}_T := \boldsymbol{x}(T) &= (\boldsymbol{V}e^{\boldsymbol{D}T} - e^{-\boldsymbol{D}T}\boldsymbol{V})\boldsymbol{c}^* + \boldsymbol{\mu} = \operatorname{Cov}(\boldsymbol{X}^1(T))e^{\boldsymbol{D}T}\frac{e^{-\boldsymbol{D}T}\boldsymbol{v}}{\boldsymbol{v}^\top \operatorname{Cov}(\boldsymbol{X}^1(T))\boldsymbol{v}}a + \boldsymbol{\mu} \\ &= \frac{\operatorname{Cov}(\boldsymbol{X}^1(T))\boldsymbol{v}}{\boldsymbol{v}^\top \operatorname{Cov}(\boldsymbol{X}^1(T))\boldsymbol{v}}\left(P_{\max} - \boldsymbol{v}^\top\boldsymbol{\mu}\right) + \boldsymbol{\mu}.\end{aligned}$$

According to Approximation 6, the most likely path from $\boldsymbol{x}_s$ at any time $s \in (0, T]$ will hit the rare event set at $\boldsymbol{x}_T$. By imposing the boundary condition $\boldsymbol{x}(T) = \boldsymbol{x}_T$ to (20), we derive the corresponding approximation $\tilde{\boldsymbol{x}}(t)$ for the most likely path

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\boldsymbol{F}^{-1}(T)\left(\boldsymbol{x}_T - \boldsymbol{G}(T)\right) + \boldsymbol{G}(t),$$

with $\boldsymbol{F}(t) := \boldsymbol{V}e^{\boldsymbol{D}t} - e^{-\boldsymbol{D}(t-s)}\boldsymbol{V}e^{\boldsymbol{D}s}$ and $\boldsymbol{G}(t) := e^{-\boldsymbol{D}(t-s)}(\boldsymbol{x}_s - \boldsymbol{\mu}) + \boldsymbol{\mu}$. This approximation avoids the optimization problem (21)-(22) too as $\boldsymbol{c}$ is completely

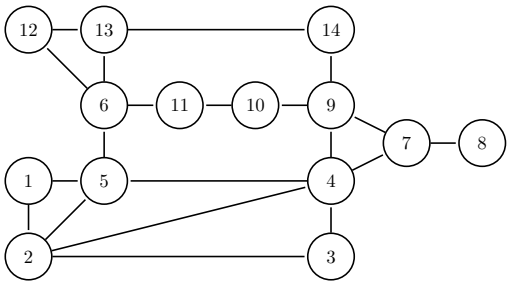

Figure 3: The topology of the IEEE-14 test network, consisting of 14 nodes. Power is consumed at nodes 2, 3, 4, 5, 6, 9, 10, 11, 12, 13 and 14. Power is generated at node 2. Node 1 is the so-called the slack node, absorbing any power imbalance in the grid, whereby its power injection does not influence the power flowing through any connection. So only nodes 2, 3, 4, 5, 6, 9, 10, 11, 12, 13 and 14 have nonzero power injections in the DC model.

determined by the entrance point $\boldsymbol{x}_T$. The corresponding approximate decay rate is

$$I_3^*(s, \boldsymbol{x}_s) = \boldsymbol{w}^\top \operatorname{Cov}(\boldsymbol{X}^1(T-s))^{-1}\boldsymbol{w}, \tag{28}$$

with $\boldsymbol{w} := \boldsymbol{x}_T - \boldsymbol{\mu} - e^{-\boldsymbol{D}(T-s)}(\boldsymbol{x}_s - \boldsymbol{\mu})$. The covariance matrix inverses are independent of the chain state $\boldsymbol{x}_s$ so they can be computed for each s before the simulation starts. $\hat{\gamma}_3$ denotes the corresponding splitting estimator.

6 Experiments

We perform experiments on the IEEE-14 test network, representing a portion of the American Electric Power System (in the Midwestern US) [Christie, 2006]. The grid contains of 14 grid nodes, 20 connections, and the original network has constant and deterministic power injections, which are nonzero at 11 nodes (see Figure 3).

At the first n nodes that have nonzero power injections, we replace these n deterministic power injections $\boldsymbol{P}^{\text{det}}$ by OU model (1) with $\boldsymbol{\mu} = \boldsymbol{x}_0 = \boldsymbol{P}^{\text{det}}$. In this way the processes tend to revert to the original deterministic power injection values. The network data assumes the per-unit system, which is widely used in the power system industry [Grainger and Stevenson, 1994] and expresses voltages, currents, powers, and impedances in per-unit (p.u.) values by dividing each by a corresponding base value. For example, since the MVA base value is 100 in the IEEE network, $P_2^{\text{det}} = 0.183$ p.u. means that 18.3 MVA is injected at node 2. By setting the end time $T = 1$ hour and step size $\Delta = 0.01$ hour, the experiment corresponds to an operational assessment to determine whether the grid is sufficiently reliable during the coming hour.

We set rarity parameter $\varepsilon = 0.1$ and the mean-reverting terms $\theta_i = 1 + (i-1)/(n-1)$ increase from 1 to 2 per hour. The lower Cholesky factor $\boldsymbol{L}$ is such that covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top = \operatorname{diag}(\boldsymbol{\Sigma}^{1/2})(\rho \mathbf{1}\mathbf{1}^\top + (1-\rho)I)\operatorname{diag}(\boldsymbol{\Sigma}^{1/2})$. Here $\rho = 0.5$ reflects the typically positive correlation of power injections, $\mathbf{1} \in \mathbb{R}^n$ is a vector of ones and the volatilities $\operatorname{diag}(\boldsymbol{\Sigma}^{1/2})_{ii} = 1 + (i-1)/(n-1)$ of the

marginal OU processes increase from 1 to 2 p.u. per square root of an hour. By choosing these parameter values, the standard deviation $\sqrt{\operatorname{Var} \boldsymbol{X}_i^\varepsilon(t)}$ of each OU process increases in time from zero towards a similar order of magnitude as the mean μ_i — that is, each generation pattern will be rather volatile at end time T.

We assume DC power flow equations implying $p(\boldsymbol{x}) = \boldsymbol{v}^\top \boldsymbol{x}$ where $\boldsymbol{v}$ depends on the connection under consideration. We use the MATPOWER package in MATLAB to extract the values for $\boldsymbol{v}$ [Zimmerman et al., 2011]. For each connection, we set the maximum allowed power flow $P_{\max} = C|\boldsymbol{v}^\top \mu|$ equal to a factor $C > 1$ times the average absolute power flow through that connection. All experiments are performed on an Intel Core 2 Quad CPU Q9550 2.83GHz computer in MATLAB R2014b.

6.1 Two nodes with stochastic power injections

We choose power injections at nodes 2 and 3 to be stochastic, so $n = 2$, and we choose $C = 1.5$. The approximate decay rate and results of a Crude Monte Carlo (CMC) simulation and splitting simulations are displayed in Table 1. Each row corresponds to a connection denoted by $i \to j$, and each column contains different probability estimates of connection overloads. To distinguish between excessive power flow in opposite directions, this probability is defined as

$$\gamma = \begin{cases} \mathbb{P}\{\sup_{\tau\in(0,T]} \boldsymbol{v}^\top \boldsymbol{X}^\varepsilon(\tau) \geq C|\boldsymbol{v}^\top \boldsymbol{\mu}|\} & \text{if } i < j, \\ \mathbb{P}\{\inf_{\tau\in(0,T]} \boldsymbol{v}^\top \boldsymbol{X}^\varepsilon(\tau) \leq -C|\boldsymbol{v}^\top \boldsymbol{\mu}|\} & \text{if } i > j. \end{cases} \tag{29}$$

The second column in Table 1 contains the largest approximate overload probabilities $\tilde{\gamma}_{\mathrm{LD}} := e^{-g_{0,\boldsymbol{\mu}}(T)/\varepsilon}$ of all connections, with $g_{0,\boldsymbol{\mu}}(T)$ as in (17). The third column contains CMC estimates $\hat{\gamma}_{\mathrm{CMC}}$ (with the relative error between parentheses) obtained from one simulation using 10^6 samples. Expression (5) suggests that obtaining a reasonably sized 95% confidence interval halfwidth of $1.96\sqrt{\operatorname{Var} \hat{\gamma}_{\mathrm{CMC}}} \leq \gamma/2$ requires more than 10^{11} samples when $\gamma < 10^{-10}$. Since this would in turn require more than 546 hours of CPU time, we omitted computing CMC estimates for which $\tilde{\gamma}_{\mathrm{LD}} < 10^{-10}$. The CMC estimates show that $\hat{\gamma}_{LD}$ is reasonably accurate for those overload probabilities we can compare to a CMC estimate. As the computation of $\hat{\gamma}_{LD}$ required the evaluation of (18) only, it serves as a suitable first guess to distinguish grid connections that are exposed to significant overloading risks.

However, the accuracy of the single-point approximations $\hat{\gamma}_{LD}$ are unknown. Fortunately, relative errors of multiple splitting estimates give this insight. The fourth column contains the means of 100 splitting estimates $\hat{\gamma}_1$ using the FNS scheme explained in Section 3 with 100 hits at each level. We omitted splitting estimates for which $\tilde{\gamma}_{\mathrm{LD}} < 10^{-25}$ as knowing such small estimates will have no practical purpose. We chose the number of equidistant thresholds to be the closest integer to $-0.6275 \log \tilde{\gamma}_{\mathrm{LD}}$, following the reasoning in [Amrein and Künsch, 2011]. We estimated the relative errors of the splitting estimators by repeating the simulation 100 times. All CMC estimates except the last one agree with the corresponding splitting estimates in the sense that the 95% confidence intervals implied by the splitting estimates contain the CMC estimates. Although the last CMC estimate does not lie in the corresponding splitting confidence interval, its relative error is quite large — in fact, the 95%

confidence intervals of CMC and splitting overlap — so CMC does not serve as a good benchmark for such a small probability.

To illustrate the computational gain of a splitting technique over CMC simulation, we use equation (5) to compute the expected number of CMC samples required to obtain an accuracy comparable to that of the splitting estimates. For example, the squared relative error of $\hat{\gamma}_1$ for connection $3 \to 4$ is $0.021^2 \approx 0.00044$. Equation (5) suggests that 9.1×10^6 CMC samples will be required to achieve a squared relative error of similar size, whereas the splitting estimator required only 2.8×10^5 samples. This difference of a factor 33 is displayed in the last column of Table 1. This factor is as much as 9.8×10^6 for connection $2 \to 4$, since CMC simulation would then require the prohibitively large number of 6.8×10^{12} samples. Even though the CPU time of one trajectory is smaller when using CMC simulation (2.5×10^{-5} seconds) than when using splitting (7.9×10^{-4}, 9.6×10^{-5} and 1.2×10^{-4} seconds for the three respective splitting algorithms from Section 5.1), Table 1 illustrates that the smaller the probability, the larger the computational gain of splitting.

To our knowledge, of all models in the literature for which rare event simulation is used to estimate grid reliability due to uncertain power injections, the one in [Nykvist, 2015, Section 5.4.3-4] is closest to ours. The authors also assume OU processes and base the simulation technique on results from large deviations, but they assume AC (instead of DC) power flow equations and employ importance sampling instead of splitting. The resulting estimates illustrate that the importance sampling algorithm is asymptotically optimal. However, the average CPU time per sample path is much higher that of the splitting schemes in this paper: the nonlinearity of the AC power flow equations will require a computationally intensive numerical optimization. Our splitting technique has a similar disadvantage: for nonlinear p, each time step of each sample path optimization problem (13) requires a numerical solver, which will probably nullify the workload gain achieved by extending CMC simulation with splitting.

6.2 Eleven nodes with stochastic power injections

We increase the number of stochastic nodes to $n = 11$ and repeat the experiment. We choose $C = 20$ in (29) to again achieve a wide range of overload probabilities, see Table 2. Again $\hat{\gamma}_{LD}$ is reasonably accurate for those probabilities we can compare to a CMC estimate, confirming that it may serve as a reasonable first guess even when a high number of stochastic power injections are involved. For all seven CMC estimates the 95% confidence intervals obtained by the corresponding splitting estimates contain the CMC estimates. Again achieving the relative error of displayed splitting estimates using a CMC simulation will often require a prohibitively large number of samples.

6.3 Comparison of the three decay rate approximations

We investigate the performance of three splitting techniques, each using one of the three decay rate approximations I_1^*, I_2^* and I_3^*. As performance measures we will use the relative error and CPU time of the respective splitting estimators $\hat{\gamma}_1, \hat{\gamma}_2$ and $\hat{\gamma}_3$ of overload probability γ of connection $3 \to 4$ in the two-dimensional model as in Section 6.1. The sample mean and relative error of 100 realizations of $\hat{\gamma}_i$ are displayed in Table 3, using 100 hits at every level in all

Table 1: Estimates of highest overload probabilities γ as in (29) for connections $i \to j$ in the DC IEEE-14 test case with $n = 2$ stochastic power injections, and $\varepsilon = 0.1$. Columns contain large deviations approximations $\tilde{\gamma}_{\text{LD}}$, CMC estimates $\hat{\gamma}_{\text{CMC}}$ and the means of 100 large deviations based splitting estimates $\hat{\gamma}_1$, with relative errors of the mean between parentheses. The last column contains the expected number of CMC samples required to obtain the accuracy of the splitting estimates, divided by the number of splitting samples.

	$\tilde{\gamma}_{\text{LD}}$	$\hat{\gamma}_{\text{CMC}}$	$\hat{\gamma}_1$	$\frac{\#\text{CMC samples}}{\#\text{FNS samples}}$
$4 \to 3$	0.72	0.69 (0.000 68)	0.69 (0.0056)	1.0
$1 \to 2$	0.11	0.090 (0.0032)	0.091 (0.0097)	1.1
$2 \to 3$	0.10	0.10 (0.0030)	0.10 (0.0093)	1.2
$5 \to 4$	0.013	0.012 (0.0090)	0.013 (0.016)	2.9
$1 \to 5$	0.0018	0.0014(0.027)	0.0015 (0.022)	7.2
$3 \to 4$	2.8E−4	2.6E−4(0.062)	2.5E−4 (0.021)	33
$11 \to 10$	7.6E−5	5.3E−5(0.14)	6.1E−5 (0.026)	73
$2 \to 4$	8.5E−11	-	6.1E−11 (0.049)	9.8E6
$9 \to 10$	6.7E−14	-	3.3E−14 (0.051)	1.2E10
$6 \to 11$	1.2E−18	-	5.1E−19 (0.083)	2.1E14
$2 \to 5$	4.0E−23	-	1.0E−23 (0.11)	5.4E18
$2 \to 1$	1.6E−24	-	3.7E−25 (0.13)	9.2E19
$3 \to 2$	1.8E−25	-	7.5E−28 (0.13)	4.5E20
$13 \to 14$	8.8E−26	-	-	

Table 2: As in Table 1, but now with $n = 11$ stochastic power injections, and $C = 20$.

	$\tilde{\gamma}_{\text{LD}}$	$\hat{\gamma}_{\text{CMC}}$	$\hat{\gamma}_1$	$\frac{\#\text{CMC samples}}{\#\text{FNS samples}}$
$12 \to 13$	0.049	0.043 (0.0047)	0.043 (0.011)	1.8
$13 \to 12$	0.025	0.021 (0.0068)	0.022 (0.012)	2.4
$9 \to 10$	0.0060	0.0045 (0.015)	0.0046 (0.019)	3.2
$10 \to 9$	0.0019	0.0014 (0.027)	0.0014 (0.021)	7.2
$11 \to 10$	6.4E−4	4.9E−4 (0.045	5.0E−4 (0.019)	24
$10 \to 11$	1.2E−4	9.3E−5 (0.10)	9.1E−5 (0.026)	56
$9 \to 14$	3.1E−11	-	1.9E−11 (0.041)	4.2E7
$6 \to 12$	1.3E−11	-	6.3E−12 (0.043)	9.6E7
$6 \to 11$	5.2E−12	-	2.2E−12 (0.048)	2.3E8
$14 \to 9$	1.5E−13	-	8.3E−14 (0.069)	2.8E9
$12 \to 6$	4.9E−14	-	2.1E−14 (0.056)	1.5E10
$11 \to 6$	1.6E−14	-	6.0E−15 (0.063)	4.1E10
$8 \to 7$	1.1E−15	-	3.1E−16 (0.060)	7.8E11
$13 \to 14$	5.6E−17	-	2.7E−17 (0.085)	4.2E12
$7 \to 8$	5.4E−19	-	1.3E−19 (0.062)	1.5E15
$14 \to 13$	1.4E−20	-	5.5E−21 (0.086)	1.7E16
$5 \to 6$	8.7E−25	-	2.5E−25 (0.11)	2.0E20
$6 \to 13$	3.2E−25	-	1.2E−25 (0.17)	1.6E20
$7 \to 9$	3.9E−29	-	-	

Table 3: Sample mean and its relative error of 100 splitting estimates $\hat{\gamma}_i$ using corresponding decay rate proxy I_i^*, the total time to compute 100 estimates, and the required CPU time to obtain the relative error of $\hat{\gamma}_1$ by computing more than 100 estimates.

	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
Sample mean	2.47E−4	2.37E−4	2.40E−4
Relative error (re)	0.019	0.033	0.037
CPU time (sec)	270	56	71
CPU time to obtain re($\hat{\gamma}_1$) (sec)	270	184	283

splitting runs. All 95% confidence intervals implied by the splitting estimates contain the corresponding CMC estimate — which is 2.6×10^{-4}, see Table 1. The relative error of $\hat{\gamma}_1$ is smaller than that of $\hat{\gamma}_2$, which is as expected as Approximation 4 is based on a weaker assumption than that of Approximation 5 (see Section 5.1). For a similar reason, the difference in relative error between $\hat{\gamma}_2$ and $\hat{\gamma}_3$ is as expected. Fortunately, the increase of the relative error is at most a factor of 2.

The higher workload of $\hat{\gamma}_1$ is obviously due to the necessary computation of good rate function (23) for all discrete candidates for τ^*. The workload of $\hat{\gamma}_2$ is smaller than that of $\hat{\gamma}_3$: the most demanding step to compute $\hat{\gamma}_3$ is the quadratic product $\boldsymbol{w}^\top \operatorname{Cov}(\boldsymbol{X}^1(T-s))^{-1}\boldsymbol{w}$ in (28), whereas to compute $\hat{\gamma}_2$, the quadratic product $\boldsymbol{v}^\top \operatorname{Cov}(\boldsymbol{X}^1(T))\boldsymbol{v}$ in (18) is most demanding. The matrix inverse of the former vector matrix vector computation depends on the time step, explaining the slightly higher workload of $\hat{\gamma}_3$. The last row in Table 3 gives an indication of the relative workload gain of the three splitting techniques: by computing $100\text{re}(\hat{\gamma}_i)^2/\text{re}(\hat{\gamma}_1)^2$ (instead of 100) estimates $\hat{\gamma}_i$, one may expect the relative error of the mean to be comparable to re($\hat{\gamma}_1$). The last row in the table displays the expected total CPU times these computations would require, favoring the second splitting technique over the other two in this example.

We will illustrate the accuracy of Approximation 6 of a constant optimal hitting time $\tau^* = T$ and constant optimal endpoint $\boldsymbol{x}_{\tau^*} = \boldsymbol{x}_T$. We consider connection $2 \to 4$ and again assume the model as in Section 6.1. For each entrance state at each level in the I_1^*-based splitting run, we numerically compute the hitting time τ^* and endpoint $\boldsymbol{x}_{\tau^*}$ that are most likely given that entrance state. We use fifteen levels and 1000 hits per level. Figure 4 displays the histograms of τ^* for four different levels. The optimal hitting time τ^* is relatively close to T for second level entrance states. This can be interpreted as many entrance states still being close to $\boldsymbol{\mu}$ — we have proven in Section 4.2 that starting from $\boldsymbol{\mu}$ the most likely hitting time is exactly T. For higher levels, typical values for τ^* decrease. This is intuitive since at higher levels samples are more likely to be so close to the rare event set that a rare event occurrence is more likely before the end time than at the end time. The increased relative error of $\hat{\gamma}_2$ and $\hat{\gamma}_3$ in Table 3 compared to the relative error of $\hat{\gamma}_1$ can be attributed to this intuition.

We perform a similar analysis on the optimal hitting point $\boldsymbol{x}_{\tau^*}$. Since the two elements of $\boldsymbol{x}_{\tau^*}$ fulfil the linear equation of the rare event, we will only investigate the first element $x_{\tau^*,1}$. Figure 5 displays the histogram of $x_{\tau^*,1}$ again for four different levels. As expected, all four empirical distributions

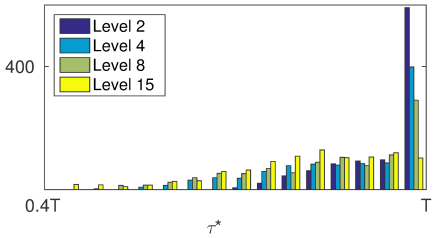

Figure 4: The optimal hitting time τ^* given an entrance state at a certain level typically decreases in this level.

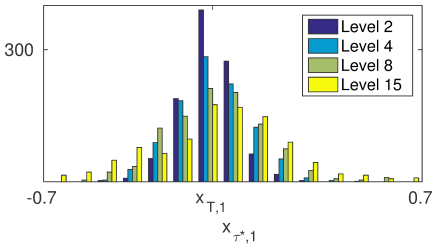

Figure 5: The endpoint $\boldsymbol{x}_{\tau^*}$ that is most likely given an entrance state at a certain level typically diverges from the endpoint $\boldsymbol{x}_T$ that was initially the most likely.

are centered around the endpoint $x_{T,1}$ that was initially the most likely. The variance increases in the considered level; in fact, the sample variance increases monotonically over all 15 levels. We can attribute this observation to the fact that the variance of $X_i^\varepsilon(t)$ increases in time (see equation (32) in Appendix A). Therefore, chain states from which the most likely endpoint is far away from $x_{T,1}$ become more likely over time.

6.4 Performance comparison with a naive importance function

Instead of employing large deviation theory, one could base an importance function on the Euclidean distance of the constraint state variable to its allowed maximum. For example, the importance function

$$h_{\mathrm{Ed}}(\boldsymbol{x}) = \frac{\boldsymbol{v}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{x}_0}{P_{\max} - \boldsymbol{v}^\top \boldsymbol{x}_0} \tag{30}$$

is zero at $\boldsymbol{x} = \boldsymbol{x}_0$ and larger than one if the rare event set is entered.

Although this choice for the importance function is intuitive, we will show in an experiment that the choice is naive since it replicates relatively unpromising sample paths: loosely speaking, a state variable $\boldsymbol{X}^\varepsilon(t)$ being closer to the rare event set does not necessarily mean that entering the rare event set is more likely. For example, if from a certain state $(t, \boldsymbol{X}^\varepsilon(t))$ only a small increase of one process $X_i(t)$ at that time t will cause the rare event, importance function (30) may duplicate sample paths from that state. However, if OU process X_i has an extremely small volatility and $X_i(t) \gg \mu_i$, then the rare event may in fact be much less likely from $(t, \boldsymbol{X}^\varepsilon(t))$ than from other realized states. Especially for problems with a high-dimensional state space $(t, \boldsymbol{X}^\varepsilon(t))$, it will be difficult to exclude the existence of such states (that are close to the rare event set in a Euclidean sense but far in a probabilistic sense). Therefore, the more grid nodes have uncertain power injections patterns, the more pronounced this problem will be.

We will illustrate that already in case of a two-dimensional $\boldsymbol{X}^\varepsilon(t)$ this problem can be significant. We call function (30) the naive importance function and compare it with importance function (24) based on decay rate approximation

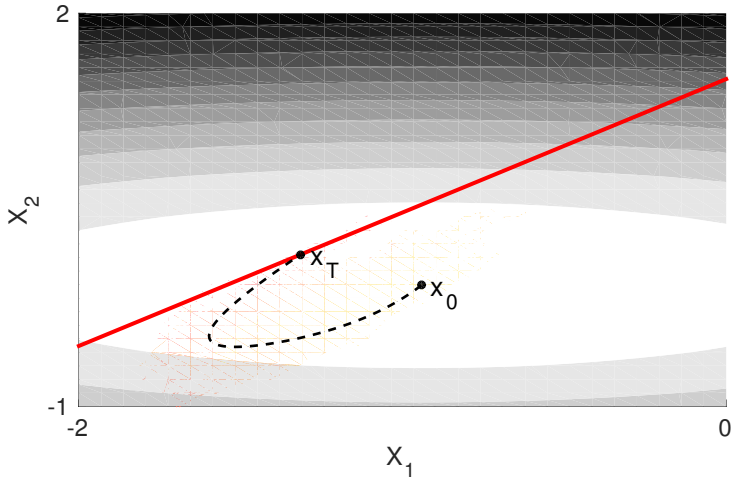

Figure 6: The most likely path of a two-dimensional OU process from $\boldsymbol{x}_0 = \boldsymbol{\mu}$ (dashed line) towards the rare event set (boundary given by solid line). A contour plot of the OU potential $\boldsymbol{D}(\boldsymbol{\mu} - \boldsymbol{x})^2/2$ is in the background. The OU model parameters are such that the most likely path differs significantly from the shortest path to the rare event.

I_2^*. We choose the model with 2 stochastic power injections as in Section 6.1, but now it is nodes 3 and 5 that are stochastic and we set $C = 1.3$, $\rho = 0.95$, $D_{22} = \theta_2 = 5$. Figure 6 shows the most likely path from $\boldsymbol{x}_0 = \boldsymbol{\mu}$ to the rare event. Because now correlation ρ is very high, path increments diagonally to the upper right and left down are much more likely than diagonally to the upper left or right down. Second, since $\theta_2 > \theta_1$, mean reversion of horizontal increments is less powerful than that of vertical increments. For these two reasons the most likely path differs significantly from the shortest path to the rare event set.

The sample paths hitting a next level are displayed in Figures 7 and 8 for the two splitting simulations, respectively. We chose only 10 hits per level for clarity reasons. Paths of the large deviations based splitting run stay around the path that is initially the most likely, whereas paths of the naive splitting run deviate to the upper right. This suggests that naive splitting is replicating many paths that are not necessarily promising to hit the rare event set. A CMC simulation with 10^7 samples yielded the estimate 3.83×10^{-5} and 95% confidence interval $[3.45 \times 10^{-5}, 4.21 \times 10^{-5}]$. The large deviation based splitting estimates in Table 4 are relatively close to the CMC estimate compared to the naive splitting estimates. Second, the relative error of 100 large deviations based estimates is lower than that of 100 naive splitting estimates, confirming that a significant number of samples are replicated in vain in the naive splitting run. This can be explained by the next statistic in Table 4: on average much more samples are required to observe a next level hit, and this difference increases for smaller r. The CPU times exhibit a similar difference. Both differences are intuitive since for a small number of next level hits in the naive run chances are higher that none of them is actually promising from a large deviation perspective. In contrast, the large deviation based splitting simulation requires around 6

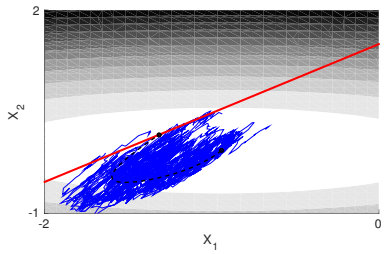

Figure 7: Using importance function (24), with decay rate proxy I_2^*, paths of a splitting simulation stay around the most likely path (see also Figure 6)

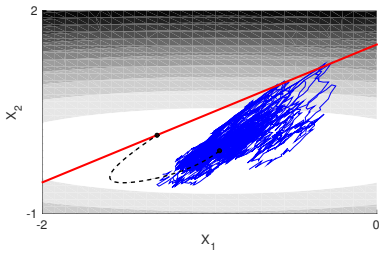

Figure 8: Using importance function (30), based on the proximity to the rare event set, paths of a splitting simulation deviate significantly from the most likely path (see also Figure 6).

Table 4: Performance statistics of 100 splitting estimates using either large deviations based importance function (24) or naive importance function (30), for different numbers r of hits per level: sample mean, relative error, average number of samples required to hit the next level and total CPU time for the 1000 estimates. The last column displays the factor of CPU time naive splitting would require more than LD based splitting to obtain the accuracy of the LD based splitting estimate.

	h_{LD}				h_{Ed}				$\frac{\text{CPU Ed}}{\text{CPU LD}}$
r	Est	RE	#paths	CPU (sec)	Est	RE	#paths	CPU (sec)	
250	3.94E−5	0.021	5.5	170	3.94E−5	0.13	13	270	60.8
100	3.85E−5	0.040	5.5	70	3.70E−5	0.26	20	160	96.5
25	3.75E−5	0.064	5.6	21	1.53E−5	0.36	85	110	166
10	3.81E−5	0.12	5.7	11	3.28E−6	0.40	707	175	177

samples on average to hit the next level, even for a small number of hits per level. In this sense, the workload of $\hat{\gamma}_2$ per level hit is robust in the number of hits per level.

To give a quantification of the workload gain, first note that for $r = 250$ the relative error is a factor $0.13/0.021 = 6.19$ smaller using h_{LD} instead of h_{Ed}. Again using expression (5), we expect the naive splitting simulation to require a factor $6.19^2 \approx 38$ as many estimates to achieve a relative error similar to that of the large deviation based splitting simulation. This would translate in a total CPU time of $270 \times 38 = 10260$ seconds, which is $10260/170 = 60.8$ times as much as the CPU time of the large deviations based splitting simulation. The last column in Table 4 shows that for $r = 100, 25, 10$ this factor of increased CPU time becomes even more. We conclude that for a fixed accuracy the large deviation based splitting technique is computationally more efficient than the naive splitting technique, especially for a relatively small number of hits per level.

Using expression (5) and an estimated CPU time of 2.5×10^{-5} seconds per CMC sample, a similar comparison of naive splitting with CMC can be performed. In fact, for $r = 250, 100, 25, 10$, the naive splitting technique required

$11, 25, 33, 67$, respectively, times as much CPU time to obtain the relative errors $0.13, 0.26, 0.36, 0.40$ in Table 4, respectively, than CMC simulation would require. So in this case CMC simulation is computationally more efficient than naive splitting. In contrast, large deviation based splitting outperforms CMC for $r = 250, 100, 25, 10$ with factors $5.7, 3.8, 5.0, 2.7$, respectively, in computational efficiency. This efficiency gain will be even more for smaller probabilities.

7 Conclusion and outlook

Based on results from large deviations theory, we developed an importance function for a splitting technique to efficiently estimate overload probabilities of power grid connections. The large deviations approximation serves as a suitable first guess to distinguish connections with significant overload probabilities. For both 2 and 11 stochastic power injections and a fixed accuracy, Crude Monte Carlo would require tens to millions as many samples than the proposed splitting technique required. The assumption that the rare event is most likely at the end time significantly accelerates the computation with only a modest loss of accuracy.

We showed an example (see Table 4) where a naive importance function based on the Euclidean distance to the rare event set replicates many unpromising sample paths. When using a naive importance function the required CPU time to achieve a fixed relative error is over 60 times larger than when using our proposed importance function, justifying the use of a large deviations based splitting technique. In fact, naive splitting — unlike large deviations based splitting — required more (over 11 times as much) CPU time than CMC simulation, illustrating its pitfall.

The splitting techniques in this paper consider overload probabilities of single connections only. The event that any connection will overload in due time will be of great importance for power system engineers. Generalizing the large deviation based splitting technique to this problem will therefore be relevant further research. Since the minimum good rate function of any connection overloading is equal to the minimum over all minimum good rate functions of a specific connection overloading, a similar large deviation based importance function can be derived for this problem. We expect that computing the minimum good rate function of each (instead of one) connection at each time step in the simulation will constitute the main increase in workload. Especially for grids with a large number of connections this workload increase may challenge the workload gain due to splitting. Estimating conditional probabilities given the overload of one line that one other line fails will also be relevant but a challenging extension of the approach in this paper: when one line overloads the change of grid topology will probably hinder analytically solving the quadratic programming problem in (4.3).

A second area of further research is to develop large deviation based importance functions for nonlinear power flow equations. In that case optimization problem (21)-(22) is multidimensional and has a nonlinear constraint, so solving it each time step will be computationally too intensive for a high dimensional state space. However, a remedy could be to use one of the three importance functions from Section 5.1 (that assume linear power flow equations) in a splitting simulation of a nonlinear power flow model. Importance function evaluations

will in this way avoid the optimization problem and may be sufficiently accurate to save a substantial amount of workload for a fixed relative error.

Finally, we aim to replace the OU process by a diffusion process that incorporates periodicities or an alternative stationary distribution that is typical for generation patterns of renewable energy.

APPENDIX

A Moments of the multidimensional OU process

In this appendix we will derive the first two moments of the multidimensional OU process. Consider the i-th element of the multidimensional OU process in (1):

$$dX_i^\varepsilon(t) = \theta_i(\mu_i - X_i^\varepsilon(t))dt + \sqrt{\varepsilon}\sum_{k=1}^{i} L_{ik}dW_k(t), \quad X_i^\varepsilon(0) = x_{0,i}.$$

Then marginal $X_i^\varepsilon(t)$ is a one dimensional OU process with volatility parameter σ_i given by $\sigma_i^2 = \varepsilon\sum_{k=1}^{i} L_{ik}^2 = \varepsilon\Sigma_{ii}$, and its solution is well-known:

$$X_i^\varepsilon(t) = x_{0,i}e^{-\theta_i t} + \mu_i(1 - e^{-\theta_i t}) + \varepsilon\sum_{k=1}^{i} L_{ik}\int_0^t e^{\theta_i(s-t)}dW_k(s).$$

The first two RHS terms are deterministic and the third is a weighted sum of independent Itō integrals with a deterministic integrand. Therefore, all Itō integrals are normally distributed with zero mean and a variance equal to the time integral of the squared integrand, implying

$$\mathbb{E}[X_i^\varepsilon(t)] = x_{0,i}e^{-\theta_i t} + \mu_i(1 - e^{-\theta_i t}), \tag{31}$$

$$\mathrm{Var}(X_i^\varepsilon(t)) = \varepsilon\sum_{k=1}^{i} L_{ik}^2\,\mathrm{Var}\int_0^t e^{\theta_i(s-t)}dW_k(s) = \varepsilon\sum_{k=1}^{i} L_{ik}^2\int_0^t e^{2\theta_i(s-t)}ds = \varepsilon\frac{\sigma_i^2}{2\theta_i}(1 - e^{-2\theta_i t}).$$

As every linear combination of components of $\boldsymbol{X}^\varepsilon(t)$ is univariate normally distributed, $\boldsymbol{X}^\varepsilon(t)$ is multivariate normal. Its expectation is the vector of above marginal expectations, so it remains to find the covariance matrix of $\boldsymbol{X}^\varepsilon(t)$. Assuming $i \leq j$ without loss of generality, the elements of this covariance matrix are

$$\begin{aligned}
\mathrm{Cov}(X_i^\varepsilon(t), X_j^\varepsilon(t)) &= \mathbb{E}[(X_i^\varepsilon(t) - \mathbb{E}[X_i^\varepsilon(t)])(X_j^\varepsilon(t) - \mathbb{E}[X_j^\varepsilon(t)])] \\
&= \varepsilon\sum_{k=1}^{i}\sum_{l=1}^{j} L_{ik}L_{jl}e^{-(\theta_i+\theta_j)t}\mathbb{E}\left[\int_0^t e^{\theta_i s}dW_k(s)\int_0^t e^{\theta_j s}dW_l(s)\right] \\
&= \varepsilon\sum_{k=1}^{i} L_{ik}L_{jk}e^{-(\theta_i+\theta_j)t}\mathbb{E}\left[\int_0^t e^{\theta_i s}dW_k(s)\int_0^t e^{\theta_j s}dW_k(s)\right],
\end{aligned}$$

where the last equality holds as for $k \neq l$ the two Itō integrals are independent and have a deterministic integrand, so the expectation of their product is zero.

The argument of the remaining expectations are quadratic covariations, of which each can be written in terms of three quadratic variations using the polarization identity:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^t e^{\theta_i s} dW_k(s) \int_0^t e^{\theta_j s} dW_k(s)\right] \\
&= \frac{1}{2}\mathbb{E}\left[\left(\int_0^t (e^{\theta_i s} + e^{\theta_j s}) dW_k(s)\right)^2 - \left(\int_0^t e^{\theta_i s} dW_k(s)\right)^2 - \left(\int_0^t e^{\theta_j s} dW_k(s)\right)^2\right] \\
&= \frac{1}{2}\left(\int_0^t (e^{\theta_i s} + e^{\theta_j s})^2 ds - \int_0^t e^{2\theta_i s} ds - \int_0^t e^{2\theta_j s} ds\right) \\
&= \int_0^t e^{(\theta_i+\theta_j)s} ds = \frac{e^{(\theta_i+\theta_j)t} - 1}{\theta_i + \theta_j}.
\end{aligned}$$

After substitution we conclude that element (i, j) of the covariance matrix of $\boldsymbol{X}^\varepsilon(t)$ is given by

$$\mathrm{Cov}(X_i^\varepsilon(t), X_j^\varepsilon(t)) = \varepsilon \sum_{k=1}^{i} L_{ik} L_{jk} e^{-(\theta_i+\theta_j)t} \frac{e^{(\theta_i+\theta_j)t} - 1}{\theta_i + \theta_j} = \varepsilon \Sigma_{ij} \frac{1 - e^{-(\theta_i+\theta_j)t}}{\theta_i + \theta_j}.$$

With $\boldsymbol{V}$ as defined in (12), we conclude in vector notation:

$$\mathrm{Cov}(\boldsymbol{X}^\varepsilon(t)) = \varepsilon(\boldsymbol{V} - e^{-\boldsymbol{D}t}\boldsymbol{V}e^{-\boldsymbol{D}t}). \tag{32}$$

B Proof of proposition 2

Necessary conditions for $\boldsymbol{x}$ are the Euler-Lagrange equations, which are in vector form as follows:

$$\nabla_{\boldsymbol{x}}\mathcal{L} - \frac{d}{dt}\left(\nabla_{\boldsymbol{x}'}\mathcal{L}\right) = 0, \tag{33}$$

where $\nabla_{\boldsymbol{x}}\mathcal{L}$ and $\nabla_{\boldsymbol{x}'}\mathcal{L}$ are the gradients of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{x}')$ w.r.t. $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. Elementary calculus yields

$$\begin{aligned}
\nabla_{\boldsymbol{x}}\mathcal{L} &= 2(\nabla_{\boldsymbol{x}}\boldsymbol{u})^\top \boldsymbol{u} = 2(\boldsymbol{L}^{-1}\boldsymbol{D})^\top \boldsymbol{L}^{-1}(\boldsymbol{x}' + \boldsymbol{D}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\mu}) = 2\boldsymbol{D}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}' + \boldsymbol{D}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\mu}), \\
\nabla_{\boldsymbol{x}'}\mathcal{L} &= 2(\nabla_{\boldsymbol{x}'}\boldsymbol{u})^\top \boldsymbol{u} = 2\boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}(\boldsymbol{x}' + \boldsymbol{D}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\mu}) = 2\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}' + \boldsymbol{D}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\mu}), \\
\frac{d}{dt}\left(\nabla_{\boldsymbol{x}'}\mathcal{L}\right) &= 2\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}'' + \boldsymbol{D}\boldsymbol{x}').
\end{aligned}$$

Therefore, (33) becomes after some rearrangements

$$\boldsymbol{x}'' = (\boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}^{-1} - \boldsymbol{D})\boldsymbol{x}' + \boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\boldsymbol{x} - \boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\boldsymbol{\mu}. \tag{34}$$

Equation (34) is a system of second order nonhomogeneous linear differential equations, and its homogeneous counterpart can be written as a first order system:

$$\boldsymbol{y}' = \boldsymbol{M}\boldsymbol{y}, \quad \text{with} \quad \boldsymbol{M} := \begin{pmatrix} \boldsymbol{O} & \boldsymbol{I} \\ \boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{D} & \boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}^{-1} - \boldsymbol{D} \end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \quad \boldsymbol{y} := \begin{pmatrix} \boldsymbol{x}_h \\ \boldsymbol{x}_h' \end{pmatrix}, \tag{35}$$

with zero matrix $\boldsymbol{O} \in \mathbb{R}^{n\times n}$ and identity matrix $\boldsymbol{I} \in \mathbb{R}^{n\times n}$. To find the eigenvalues and eigenvectors of $\boldsymbol{M}$, we have to solve

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{I} \\ \boldsymbol{\Sigma D \Sigma}^{-1}\boldsymbol{D} & \boldsymbol{\Sigma D \Sigma}^{-1} - \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \bar{\boldsymbol{w}} \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{w} \\ \bar{\boldsymbol{w}} \end{pmatrix}$$

for $\lambda \in \mathbb{R}, \boldsymbol{w}, \bar{\boldsymbol{w}} \in \mathbb{R}^n$. The upper block equation reads $\bar{\boldsymbol{w}} = \lambda \boldsymbol{w}$, so each eigenvector will be of the form $(\boldsymbol{w},\ \lambda_i \boldsymbol{w})^\top$. Substituting $\bar{\boldsymbol{w}} = \lambda \boldsymbol{w}$ in the lower block equation yields the characteristic polynomial

$$(\boldsymbol{\Sigma D \Sigma}^{-1} - \lambda \boldsymbol{I})(\boldsymbol{D} + \lambda \boldsymbol{I})\boldsymbol{w} = 0. \tag{36}$$

Any eigenvalue of $-\boldsymbol{D}$ for λ together with a corresponding eigenvector for $\boldsymbol{w}$ would obviously solve this equation. As $-\boldsymbol{D}$ is diagonal, these eigenvalues are $-D_{11}, \ldots, -D_{nn}$ with the standard unit vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ as corresponding eigenvectors. Therefore, for $i = 1, \ldots, n$, $\lambda_i := -D_{ii}$ is an eigenvalue of $\boldsymbol{M}$ and $(\boldsymbol{w}_i,\ \bar{\boldsymbol{w}}_i)^\top := (\boldsymbol{e}_i\ -D_{ii}\boldsymbol{e}_i)^\top$ is the corresponding eigenvector. Note that here $\boldsymbol{w}_i \in \mathbb{R}^n$ denotes the i-th eigenvector, and not the i-th element of a vector. Likewise, any eigenvalue of $\boldsymbol{\Sigma D \Sigma}^{-1}$ for λ with the corresponding eigenvector for $(\boldsymbol{D} + \lambda \boldsymbol{I})\boldsymbol{w}$ would also solve characteristic polynomial (36). As $\boldsymbol{\Sigma D \Sigma}^{-1}$ and $\boldsymbol{D}$ are similar matrices, $\boldsymbol{\Sigma D \Sigma}^{-1}$ has eigenvalue D_{ii} with eigenvector $\boldsymbol{\Sigma} \boldsymbol{e}_i$ for $i = 1, \ldots, n$. Therefore, for $i = 1, \ldots, n$, $\lambda_{n+i} := D_{ii}$ is an eigenvalue of $\boldsymbol{M}$ and $(\boldsymbol{w}_{n+i},\ \bar{\boldsymbol{w}}_{n+i})^\top := ((\boldsymbol{D}+D_{ii}\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{e}_i,\ (\boldsymbol{D}+D_{ii}\boldsymbol{I})^{-1}D_{ii}\boldsymbol{\Sigma}\boldsymbol{e}_i)^\top$ is the corresponding eigenvector. So now we have specified the general solution

$$\boldsymbol{y}(t) = \sum_{i=1}^{2n} c_i e^{\lambda_i t} \begin{pmatrix} \boldsymbol{w}_i \\ \bar{\boldsymbol{w}}_i \end{pmatrix}$$

of (35) up to constants $c_1, \ldots, c_{2n}$, since for all $i = 1, \ldots, n$

$$\begin{aligned} \lambda_i = -D_{ii}, \quad \begin{pmatrix} \boldsymbol{w}_i \\ \bar{\boldsymbol{w}}_i \end{pmatrix} &= \begin{pmatrix} \boldsymbol{e}_i \\ -D_{ii}\boldsymbol{e}_i \end{pmatrix}, \\ \lambda_{n+i} = \quad D_{ii}, \quad \begin{pmatrix} \boldsymbol{w}_{n+i} \\ \bar{\boldsymbol{w}}_{n+i} \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{D} + D_{ii}\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{e}_i \\ D_{ii}(\boldsymbol{D} + D_{ii}\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{e}_i \end{pmatrix}. \end{aligned}$$

The homogeneous solution $\boldsymbol{x}_h$ of (34) is by definition (see (35)) the subvector with the first n elements of $\boldsymbol{y}$, or in matrix-vector form

$$\boldsymbol{x}_h(t) = \begin{pmatrix} \boldsymbol{w}_1 & \ldots & \boldsymbol{w}_{2n} \end{pmatrix} \operatorname{diag}\left(e^{\lambda_1 t}, \ldots, e^{\lambda_{2n} t}\right) \begin{pmatrix} \bar{\boldsymbol{c}} \\ \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{V} \end{pmatrix} \begin{pmatrix} e^{-\boldsymbol{D}t} & \boldsymbol{O} \\ \boldsymbol{O} & e^{\boldsymbol{D}t} \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{c}} \\ \boldsymbol{c} \end{pmatrix}$$

with $\bar{\boldsymbol{c}} = (c_1, \ldots, c_n)^\top$, $\boldsymbol{c} = (c_{n+1}, \ldots, c_{2n})^\top$ and matrix $\boldsymbol{V}$ as defined in (12). Using particular solution $\boldsymbol{x}_p = \boldsymbol{\mu}$ of (34) and initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$ to determine $\bar{\boldsymbol{c}}$, we arrive at the general solution of (34):

$$\boldsymbol{x}(t) = (\boldsymbol{V}e^{\boldsymbol{D}t} - e^{-\boldsymbol{D}t}\boldsymbol{V})\boldsymbol{c} + e^{-\boldsymbol{D}t}(\boldsymbol{x}_0 - \boldsymbol{\mu}) + \boldsymbol{\mu}.$$

We conclude that the optimal path is of the above form, with $\boldsymbol{c} \in \mathbb{R}^n$ such that $p(\boldsymbol{x}(\tau)) = P_{\max}$.

References

[Amrein and Künsch, 2011] Amrein, M. and Künsch, H. R. (2011). A variant of importance splitting for rare event estimation: Fixed number of successes. *ACM Transactions on Modeling and Computer Simulation*, 21(2):13:1–13:20.

[Billinton and Li, 1994] Billinton, R. and Li, W. (1994). *Reliability assessment of electrical power systems using Monte Carlo methods*. Springer.

[Bosman et al., 2014] Bosman, J., Nair, J., and Zwart, B. (2014). On the probability of current and temperature overloading in power grids: a large deviations approach. *ACM SIGMETRICS Performance Evaluation Review*, 42(2):33–35.

[Botev and Kroese, 2012] Botev, Z. I. and Kroese, D. P. (2012). Efficient Monte Carlo simulation via the generalized splitting method. *Statistics and Computing*, 22(1):1–16.

[Carden and Wintermantel, 2013] Carden, K. and Wintermantel, N. (2013). The economic ramifications of resource adequacy. Technical report, Astrape Consulting.

[CEER, 2014] CEER (2014). On the continuity of electricity supply. Technical report, Council of European Energy Regulators. Available at `www.ceer.eu/`.

[Christie, 2006] Christie, R. (2006). Power systems test case archive. `http://www.ee.washington.edu/research/pstca/`.

[Dean and Dupuis, 2009] Dean, T. and Dupuis, P. (2009). Splitting for rare event simulation: a large deviation approach to design and analysis. *Stochastic processes and their applications*, 119(2):562–587.

[Dembo and Zeitouni, 2009] Dembo, A. and Zeitouni, O. (2009). *Large deviations techniques and applications*, volume 38. Springer Science & Business Media.

[Dupuis et al., 2012] Dupuis, P., Spiliopoulos, K., and Wang, H. (2012). Importance sampling for multiscale diffusions. *Multiscale Modeling & Simulation*, 10(1):1–27.

[Garvels, 2000] Garvels, M. J. (2000). *The splitting method in rare event simulation*. PhD thesis, University of Twente, Enschede, Netherlands.

[Garvels et al., 2002] Garvels, M. J., Van Ommeren, J.-K. C., and Kroese, D. P. (2002). On the importance function in splitting simulation. *European Transactions on Telecommunications*, 13(4):363–371.

[Glasserman et al., 1998] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. (1998). A large deviations perspective on the efficiency of multilevel splitting. *Automatic Control, IEEE Transactions on*, 43(12):1666–1679.

[Grainger and Stevenson, 1994] Grainger, J. J. and Stevenson, W. D. (1994). *Power system analysis*, volume 621. McGraw-Hill.

[Guasoni and Robertson, 2008] Guasoni, P. and Robertson, S. (2008). Optimal importance sampling with explicit formulas in continuous time. *Finance and Stochastics*, 12(1):1–19.

[L'Ecuyer et al., 2006] L'Ecuyer, P., Demers, V., and Tuffin, B. (2006). Splitting for rare-event simulation. In *Proceedings of the Winter Simulation Conference 2006*, pages 137–148. IEEE.

[Leith et al., 2004] Leith, D. J., Heidl, M., and Ringwood, J. V. (2004). Gaussian process prior models for electrical load forecasting. *Probabilistic Methods Applied to Power Systems*, pages 112–117.

[Low, 2014] Low, S. H. (2014). Convex relaxation of optimal power flow, part I: formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27.

[Miretskiy et al., 2012] Miretskiy, D. I., Scheinhardt, W. R., and Mandjes, M. R. (2012). On efficiency of multilevel splitting. *Communications in Statistics-Simulation and Computation*, 41(6):890–904.

[Murty, 2009] Murty, K. G. (2009). *Optimization for decision making*. Springer.

[Nykvist, 2015] Nykvist, J. (2015). *Topics in importance sampling and derivatives pricing*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

[Perninge et al., 2011] Perninge, M., Knazkins, V., Amelin, M., and Söder, L. (2011). Modeling the electric power consumption in a multi-area system. *European transactions on electrical power*, 21(1):413–423.

[Rubino and Tuffin, 2009] Rubino, G. and Tuffin, B. (2009). *Rare event simulation using Monte Carlo methods*. John Wiley & Sons.

[Schlapfer and Mancarella, 2011] Schlapfer, M. and Mancarella, P. (2011). Probabilistic modeling and simulation of transmission line temperatures under fluctuating power flows. *Power Delivery, IEEE Transactions on*, 26(4):2235–2243.

[Seifi and Sepasian, 2011] Seifi, H. and Sepasian, M. S. (2011). *Electric power system planning: issues, algorithms and solutions*. Springer Science & Business Media.

[Shortle, 2013] Shortle, J. F. (2013). Efficient simulation of blackout probabilities using splitting. *International Journal of Electrical Power & Energy Systems*, 44(1):743–751.

[Touchette, 2009] Touchette, H. (2009). The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69.

[Vanden-Eijnden and Weare, 2012] Vanden-Eijnden, E. and Weare, J. (2012). Rare event simulation of small noise diffusions. *Communications on Pure and Applied Mathematics*, 65(12):1770–1803.

[Wadman et al., 2012] Wadman, W. S., Bloemhof, G., Crommelin, D. T., and Frank, J. E. (2012). Probabilistic power flow simulation allowing temporary current overloading (volume to appear). In *Conference proceedings on Probabilistic Methods Applied to Power Systems*, pages 494–499.

[Wadman et al., 2013] Wadman, W. S., Crommelin, D. T., and Frank, J. E. (2013). Applying a splitting technique to estimate electrical grid reliability. In *Winter Simulation Conference proceedings*, pages 577–588.

[Wadman et al., 2014] Wadman, W. S., Crommelin, D. T., and Frank, J. E. (2014). A separated splitting technique for disconnected rare event sets. In *Winter Simulation Conference proceedings*, pages 522–532.

[Wan et al., 1999] Wan, H., McCalley, J. D., and Vittal, V. (1999). Increasing thermal rating by risk analysis. *Power Systems, IEEE Transactions on*, 14(3):815–828.

[Wang et al., 2011] Wang, S.-P., Chen, A., Liu, C.-W., Chen, C.-H., and Shortle, J. (2011). Rare-event splitting simulation for analysis of power system blackouts. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–7. IEEE.

[Zimmerman et al., 2011] Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2011). MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19.