Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Large deviations for Markov jump processes with uniformly diminishing rates

Andrea Agazzi[1], Luisa Andreis[2], Robert I. A. Patterson[2], D. R. Michiel Renger[2]

submitted: February 24, 2021

[1] Department of Mathematics
Duke University
120 Science Drive
Durham, NC 27710
USA
E-Mail: agazzi@math.duke.edu

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: luisa.andreis@wias-berlin.de
robert.i.a.patterson@wias-berlin.de
d.r.michiel.renger@wias-berlin.de

No. 2816
Berlin 2021

2020 *Mathematics Subject Classification.* 60F10, 60J76, 80A30.

Key words and phrases. Large deviations, chemical reaction networks, jump Markov processes, Freidlin–Wentzell theory.

AA was partially supported by the Swiss National Science Foundation grant P2GEP2-175015 and by the NSF grant DMS-1613337. He further acknowledges the hospitality of the Weierstrass Institute for Applied Analysis and Stochastics. LA acknowledges the hospitality of the Hausdorff Institute in Bonn during the Junior trimester program “Randomness, PDEs and Nonlinear Fluctuation”, where she partially worked on this project. RP and MR received support from the Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 “Scaling Cascades in Complex Systems”, Project C08.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Large deviations for Markov jump processes with uniformly diminishing rates

Andrea Agazzi, Luisa Andreis, Robert I. A. Patterson, D. R. Michiel Renger

Abstract

We prove a large-deviation principle (LDP) for the sample paths of jump Markov processes in the small noise limit when, possibly, *all* the jump rates vanish *uniformly*, but slowly enough, in a region of the state space. We further show that our assumptions on the decay of the jump rates are optimal. As a direct application of this work we relax the assumptions needed for the application of LDPs to, *e.g.*, Chemical Reaction Network dynamics, where vanishing reaction rates arise naturally particularly the context of Mass action kinetics.

1 Introduction

1.1 Large deviations of Markov jump processes

We study a family of d-dimensional Markov jump processes $\{X^v\}_{v\in\mathbb{N}}$ with state space $(v^{-1}\mathbb{Z})^d$, deterministic initial condition $X^v(0) = x_0^v \in (v^{-1}\mathbb{Z})^d$ and generator:

$$(\mathcal{L}^v f)(x) := v \sum_{r\in\mathcal{R}} \Lambda_r^v(x)\big[f(x + v^{-1}\gamma^r) - f(x)\big]. \tag{1.1}$$

Here $\mathcal{R}$ is the finite set of possible jumps, $\gamma^r \in \mathbb{Z}^d$ are the fixed jump vectors, and $v\Lambda_r^v : (v^{-1}\mathbb{Z})^d \to [0, \infty)$ the associated jump rates. The parameter v controls the noise in the system, and the scaling is chosen so that $\Lambda_r^v(x)$ converge as $v \to \infty$. Under this scaling it is known that the paths X^v concentrate on solutions of the *fluid limit* ODE [Kur70]:

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \sum_{r\in\mathcal{R}} \lambda_r(x(t))\gamma^r\,, \tag{1.2}$$

where the continuous rates $\lambda_r : \mathbb{R}^d \to [0, \infty)$ are the limits of $\Lambda_r^v(x)$ as $v \to \infty$.

The process (1.1) and the ODE (1.2) are used as microscopic and macroscopic models for a wide range of applications. For example, in the context of chemical reactions X^v denotes the concentrations (number of molecules per unit volume) of d species being transformed by a set of reactions $\mathcal{R}$. Here, for each reaction $r \in \mathcal{R}$, the vectors γ^r encode which species are removed and created when a reaction r occurs, and $\lambda_r(x)$ are the reaction rates [Kur72]. In that setting v can be interpreted as the volume size over which the concentrations are averaged. Other typical applications using similar models include biological systems involving predator-prey interaction, birth/cell division and death, biological fitness models, as well as epidemiological models. In these settings, large-deviation techniques are often applied to simplify the dynamical landscape of the complex, high dimensional microscopic model (1.1) while retrieving quantitative information about the random

DOI 10.20347/WIAS.PREPRINT.2816

fluctuations around the mean (1.2). This information can be used for example to study non-equilibrium thermodynamics [MPPR15], to speed up simulations of rare events [DM04, WRVE04, GVE19], or to study spontaneous transitions between metastable states [FW12].

The classical proof of the Large-Deviation Principle (LDP) uses a tilting, also called a change-of-measure technique. The main challenge there is that the tilting can only be performed around sufficiently regular paths, whereas the large-deviation principle needs to be proven for any non-typical path. Therefore, the large-deviation lower bound requires an approximation argument, either for the random process or for the rate functional. A particular challenge in either case is to approximate a path without changing its starting point, which is required when proving the large-deviation principle under a deterministic initial condition. This becomes more difficult if the jump rates vanish in some regions of the state space, which is however an inherent property of the models used in many application domains. For example, in the context of chemical kinetics, it is natural to assume that the rate of a chemical reaction vanishes when the concentration of one or more of the reactants approaches 0. Similarly in the context of infectious disease models, the rate of spread of a virus is usually modelled as a linear function of the infected population.

Example 1.1. *The problem is nicely illustrated by the simplest model for autocatalysis or cell division, in chemical notation:* $\mathsf{A} \to 2\mathsf{A}$*. In this case there is only one species and one jump, so we may write the linear jump rate as* $\Lambda^v(x) = \lambda(x) = x$*, starting from a concentration with one particle, i.e.* $X^v(0) = x_0^v = 1/v$*. Clearly, the process converges to the solution of* $\dot{x}(t) = x(t)$ *with initial condition* $x(0) = x_0 = 0$*, that is* $x(t) \equiv 0$*. In oher words, the process is expected to stick to the degenerate set* $\partial\mathcal{S} = \{x = 0\}$*, which corresponds to the boundary of the state space* $\mathbb{R}_+$*. However, it can be calculated (as an application of [Kor97, Lemma 2.1]) that* $\lim_{v\to\infty} v^{-1} \log \mathbb{P}\big(X^v(t) \geq \delta\big) = \delta \log(1 - e^{-t})$ *for any fixed time* $t > 0$ *and* $\delta > 0$*. Although this is a large-deviation result about the marginal* $X^v(t)$*, it suggests that the paths can also "escape" from the boundary with finite large-deviation cost. On the other hand, we note that choosing* $x_0^v \equiv 0$ *implies that* $X^v(t) \equiv 0$ *almost surely. Hence, whether or not escape is possible depends strongly on the initial condition. This is a similar principle as what is sometimes called a "well-prepared initial condition" in* Γ*-convergence theory [Mie16].*

For more general models of the type (1.1), one expects that as long as the jump rates $\lambda_r(x)$ do not vanish too rapidly approaching the degenerate points and when starting at a well-chosen initial condition near such points, then the process will be able to escape with finite large-deviation cost, and the large-deviation principle should still hold. In this paper, we show that this is indeed the case. More specifically, denoting by $\partial\mathcal{S}$ the set where some – or possibly *all* – of the reaction rates λ vanish, we prove that when the rates near $\partial\mathcal{S}$ decrease slower than $\exp\left(-1/\mathrm{dist}(x, \partial\mathcal{S})^\alpha\right)$, *i.e.*,

$$\mathrm{dist}(x, \partial\mathcal{S})^\alpha \log \lambda(x) \to 0 \qquad \text{as } \mathrm{dist}(x, \partial\mathcal{S}) \to 0, \tag{1.3}$$

the large-deviation principle holds under the assumption $\alpha \in [0, 1)$. Furthermore, we show that these conditions on the decay of the rates close to the degenerate set are, in some sense, necessary and not just sufficient.

1.2 Literature and approach

Early papers [Fen94, Lé95] establishing a sample-path large-deviation principle for jump Markov processes mimicked the Dawson–Gärtner approach [DG87], where one first derives an abstract large-deviation result for the empirical measure on paths, and then contracts it to obtain a large-deviation principle for the path of the empirical measure. Another approach, now considered the classic tilting technique, was first used in [SW95] assuming that the jump rates are uniformly bounded away from zero in the domain of interest. In [SW05], the authors relaxed this condition by assuming the existence of jumps with rates that are uniformly bounded away from zero and that "push" the process away from the degeneracy region. This assumption is further relaxed in [DRW16, ADE18a, ADE18b, ACKN20], assuming only the existence of a *sequence* of jumps that sequentially transport the process away from the problematic region.

Recent works have taken steps to generalise these assumptions, in the context of chemical kinetics [PR19] and in the context of infectious disease models [PSK16, PSK17]. These papers give sufficient conditions on the models at hand to bypass the technical difficulties encountered in the proof of the LDP lower bound when some of the jump rates are not bounded away from zero. We mention that the work [PR19] assumes "sufficiently random" initial conditions to bypass the problem, but we shall focus on a deterministic initial condition.

The problem of vanishing rates is addressed more completely in [PSK16, PSK17], where the authors obtain large-deviation estimates for vanishing rates when the microscopic initial condition allows escaping the boundary with positive – although vanishing in v – probability. Their approach is based on a careful adaptation of the standard tilting argument to obtain the LDP lower bound for processes. In particular, to bypass the the problem of jump rates vanishing in some regions of state space, the change of measure performed by the authors depends on the large-deviation scaling parameter v, which is inversely proportional to the jump size. This replaces the problem of escaping to an $\mathcal{O}(1)$ distance from these degeneracy regions uniformly in v to the one of escaping to $\mathcal{O}(1/\sqrt{v})$. Their result allows for jump rates that behave as in (1.3) for $\alpha \in [0, 1/2)$.

In this paper we bypass the change-of-measure technique altogether establishing more direct and concise bounds on the process level. The main challange in realizing this strategy is to identify a set of paths occurring with sufficient probability to recover the LDP lower bound estimate while allowing for simple estimation of such probability. The core of the approach consists of showing separate estimates on the total number of jumps and on the types of jumps for paths in such set. This allows us to extend the assumption (1.3) to any $\alpha \in [0, 1)$ while covering a larger family of processes than the existing literature. In addition, we provide a counterexample showing that our upper bound for the exponent α is optimal: If the rates of the process decay as (1.3) for $\alpha \geq 1$ with sufficient uniformity, as we make precise below, the process will no longer be able to escape the boundary with finite large-deviation cost.

1.3 Outline

The paper is structured as follows. In Section 2 we introduce our notation, we list our assumptions and we state our main result, Theorem 1, namely the LDP. We also illustrate the generality of our result with some examples. The proof of Theorem 1 is split into two sections: In Section 3 we prove the LDP lower bound, while Section 4 deals with the LDP upper bound. Finally, in Section 5, we discuss the optimality of our assumptions on the decay of the rates and in Section 6 we summarize the key quantities of the proof.

2 Notation and results

We start by giving a concrete example of the properties of systems we aim to generalize in this paper, introducing some important quantities in an intuitive way:

Example 2.1 (Mass action kinetics). *In the context of chemical kinetics, one indexes the dimension of state space with a set of* species $\{\mathsf{S}_i\}$ *representing the chemical compounds in the system of interest. To describe interactions between different compounds one defines reactions* $r \in \mathcal{R}$ *via* $\gamma^{r,\mathrm{in}}, \gamma^{r,\mathrm{out}} \in \mathbb{N}_0^d$ *and*

$$\sum_{i=1}^{d} \gamma_i^{r,\mathrm{in}} \mathsf{S}_i \longrightarrow \sum_{i=1}^{d} \gamma_i^{r,\mathrm{out}} \mathsf{S}_i.$$

This is to be understood as saying $\gamma_i^{r,\mathrm{in}}$ *copies of species* S_i *are consumed in each* r*-reaction while* $\gamma_i^{r,\mathrm{out}}$ *copies are produced. The reaction rate* Λ_r^v *is specified via a rate constant* $k_r \geq 0$

$$\Lambda_r^v(x) := k_r \frac{1}{v^{\sum_i |\gamma_i^{r,\mathrm{in}}|}} \prod_{i=1}^{d} \binom{vx_i}{\gamma_i^{r,\mathrm{in}}} \gamma_i^{r,\mathrm{in}}! \qquad \forall x \in \left(v^{-1}\mathbb{N}_0\right)^d.,$$

where $\binom{\cdot}{\cdot}$ *deontes the binomial coefficient. The jump vector is* $\gamma^r := \gamma^{r,\mathrm{out}} - \gamma^{r,\mathrm{in}}$*. These rates are bounded from above on compact sets, and they converge to* $\lambda_r(x) = k_r \prod_{i=1}^d x_i^{\gamma_i^{r,\mathrm{in}}}$ *as* $v \to \infty$*. It is easy to see that* $\Lambda_r^v(x) = 0$ *whenever* $vx_i < \gamma^{r,\mathrm{in}}$*, so that* $X^v \in (v^{-1}\mathbb{N}_0)^d$ *almost surely. As we shall illustrate in examples below, different choices of reactions result in* X^v *being confined on subsets of* $(v^{-1}\mathbb{N}_0)^d$*.*

We start by defining the set of reachable points of the process. Throughout, we fix a sequence of deterministic initial conditions $\{x_0^v\}$. By the potentially degenerate character of the stochastic dynamics at hand, we reduce the state space $(v^{-1}\mathbb{Z})^d$ to the set of reachable points of the process with that initial condition $x_0^v \in (v^{-1}\mathbb{Z})^d$:

$$\mathcal{S}_v := \{x \in \left(v^{-1}\mathbb{Z}\right)^d : \mathbb{P}[\exists t \geq 0,\ X^v(t) = x \mid X^v(0) = x_0^v] > 0\}.$$

Note that, by definition, $\Lambda_r^v(x) = 0$ whenever $x + v^{-1}\gamma^r \notin \mathcal{S}_v$ for any $x \in \mathcal{S}_v$. Assuming that in the limit $v \to \infty$, the initial values $x_0^v \in \left(v^{-1}\mathbb{Z}\right)^d$ converge to $x_0 \in \mathbb{R}^d$, we write $\mathcal{S} = \bigcap_{n\in\mathbb{N}} \overline{\bigcup_{v\geq n} \mathcal{S}_v}$ where the raised line indicates topological closure. We assume throughout that $\mathcal{S}$ is compact, but discuss how to relax this assumption in Remark 2.6. We associate to $\mathcal{S}$ the set of jumps

$$\mathcal{R}_{\geq 0} := \{r \in \mathcal{R} \colon \exists x \in \mathcal{S},\ \lambda_r(x) > 0\}$$

Notice that, depending on the sequence of initial conditions, the same Markov process may have a different state space $\mathcal{S}_v$ and different set of jumps $\mathcal{R}_{\geq 0}$. We refer to Example 2.3 for a situation where $\mathcal{R}_{\geq 0} \neq \mathcal{R}$. However, by abuse of notation, we will drop the index ≥ 0 and refer to this set simply as $\mathcal{R}$.

Finally, we define the *degenerate* set – also referred to as “boundary” from its topological characterization in many application domains – as $\partial\mathcal{S} := \{x \in \mathcal{S} : \exists r \in \mathcal{R},\ \lambda_r(x) = 0\}$. This represents the set of points where the limiting process is degenerate, *i.e.*, where the classical proof of the large-deviation principle will not immediately apply. Observe that this is a slight abuse of both notation and terminology, since this degenerate set $\partial\mathcal{S}$ may be different from the actual topological boundary of the set $\mathcal{S}$.

The following example clarifies the role of the sequence of initial conditions on the resulting LDP.

Example 2.2. *The mass action kinetics model* A $\leftrightarrow$ B *(see Example 2.1 for definition of the rates and jump vectors) with initial conditions* $x_0^v = (0, 1 + 1/v)$ *results in* $\mathcal{S}_v = \{x \in (v^{-1}\mathbb{Z}_{\geq 0})^2 \;:\; x_1 + x_2 = 1 + 1/v\}$ *and* $\mathcal{S} = \{x \in \mathbb{R}^2_{\geq 0} \;:\; x_1 + x_2 = 1\}$.

Example 2.3. *The nontrivial effect of different sequences of initial conditions is captured by the system*

$$\mathsf{B} \leftrightarrow 2\mathsf{B} \qquad \mathsf{A} \leftrightarrow \mathsf{A} + \mathsf{B},$$

with mass action kinetics (see Example 2.1). For this model, the sequence $x_0^v = (1/v, 0)$ *results in* $\mathcal{S}_v = (v^{-1}\mathbb{Z}_{\geq 0})^2 \setminus \{0\}$ *and* $\mathcal{S} = \mathbb{R}^2_{\geq 0}$*. However, if* $x_0^v = (0, 1/v)$ *we have* $\mathcal{R}_{\geq 0} = \{b \to 2B, 2B \to B\}$ *and the dynamics are restricted to* $\mathcal{S}_v = \{0\} \times (v^{-1}\mathbb{Z}_{\geq 0})$ *resulting in* $\mathcal{S} = \{x \in \mathbb{R}^2_{\geq 0} \;:\; x_1 = 0\}$.

2.1 Assumptions

To ensure existence of the limit, we require the reaction rates to satisfy some conditions.

Assumption 1 (Convergence and regularity of rates)**.** *We assume the following.*

a) There exists a collection of non-negative functions $\{\lambda_r\}_{r\in\mathcal{R}}$*, Lipschitz continuous on a neighborhood of* $\mathcal{S}$ *in* $\mathbb{R}^d$*, such that*

$$\lim_{v\to\infty} \sup_{x\in\mathcal{S}_v} \sum_{r\in\mathcal{R}} |\Lambda_r^v(x) - \lambda_r(x)| = 0.$$

b) There exists $\aleph > 0$ *such that for all* $r \in \mathcal{R}$*,* $v > 0$ *and* $x \in \mathcal{S}_v$ *with* $\Lambda_r^v(x) > 0$*, we have*

$$\frac{\Lambda_r^v(x)}{\lambda_r(x)} \geq \aleph.$$

As we ouline in Section 3.1 the proof of our main theorem is based on the construction of short linear paths moving the process away from the boundary $\partial\mathcal{S}$. We now introduce notation to decompose the state space into subsets, in each of which the linear path will be fixed. More precisely, following a standard approach first presented in [SW05], we cover the state space $\mathcal{S}$ with the relative interior of countably many convex, compact sets $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$ with $\mathcal{A}_i \subseteq \mathcal{S}$ for all $i \in \mathcal{I}$. We then define $\partial\mathcal{A}_i := \partial\mathcal{S} \cap \mathcal{A}_i$ and let $\mathcal{I}^{\mathrm{bd}} \subseteq \mathcal{I}$ be the subset of indices for which $\partial\mathcal{A}_i \neq \varnothing$.

We assume that, whenever the process starts from an initial condition close to $\partial\mathcal{S}$ (where possibly all the rates are zero), one can identify a finite sequence of favorable jumps, which we call the *escape sequence*, that push the process away from the boundary. We further crucially assume that the rates of such favorable jumps do not decay too fast as we approach the boundary. This is captured by the following example.

Example 2.4. *Consider the family of Markov jump processes* $\{X^v\}$ *with generator*

$$\mathcal{L}^v f(x) := v e^{-\frac{k}{x}} \left(f(x + v^{-1}) - f(x) \right) \qquad \text{for } f \;:\; v^{-1}\mathbb{N}_0 \to \mathbb{R}\,, \tag{2.1}$$

for any $k > 0$*. The above process, which for small* x *is a time-changed version of the autocatalytic process introduced in Example 1.1, has only one possible jump in the positive direction with rate* $\Lambda^v(x) = e^{-\frac{k}{x}}$ *s.t.* $\lim_{\varrho\to 0} \varrho \left(\inf_{x:\, x\geq\varrho} \log \lambda(x)\right) = -k \neq 0$ *. For the sequence of initial conditions* $x_0^v = 1/v$*, we have*

$\mathcal{S}_v = v^{-1}\mathbb{N}$ and $\mathcal{S} = \mathbb{R}_{\geq 0}$. Then, for any $w > 0$ and $\varepsilon \in (0, w/2)$ the probability of observing a realization of X^v in an ε-neighborhood of the path $z(s) = sw$ on the interval $s \in [0,1]$ can be trivially estimated as

$$\mathbb{P}\left[\sup_{t\in[0,1]} |X_t^v - z(t)| \leq \varepsilon\right] \leq \mathbb{P}[X_1^v \geq w - \varepsilon] \leq \mathbb{P}[X_1^v \geq \widetilde{w}]$$

for $\widetilde{w} = \min(k, w-\varepsilon)/2$. Denoting by τ_i the waiting time between the i-th and $i+1$-th jump of the Poisson process X_t^v at $x \in \mathcal{S}_v$, we further have

$$\mathbb{P}[X_1^v \geq \widetilde{w}] \leq \prod_{i=1}^{\lfloor v\widetilde{w}\rfloor} \mathbb{P}[\tau_i \leq 1] = \prod_{i=1}^{\lfloor v\widetilde{w}\rfloor} 1 - \exp[-(ve^{-kv/i})] \leq \exp\left(\sum_{i=1}^{\lfloor v\widetilde{w}\rfloor} (\log v - kv/i)\right).$$

The rough estimate above yields

$$\frac{1}{v}\log \mathbb{P}[X_1^v \geq \widetilde{w}] \leq \frac{1}{v}\lfloor v\widetilde{w}\rfloor \log v - k\sum_{i=1}^{\lfloor v\widetilde{w}\rfloor}\frac{1}{i} < (\widetilde{w} - k)\log v - k(1 + \log\widetilde{w}) \tag{2.2}$$

which approaches $-\infty$ as $v \to \infty$.

As the example above shows, sufficiently fast decay of the rates of the process X^v implies the divergence of the large-deviation cost of any nontrivial path starting on the boundary $\partial\mathcal{S}$. We now proceed to give sufficient assumptions guaranteeing that this does not happen in the general setting. In particular, to capture the idea of escaping a boundary in the higher dimensional setting, we define directions w with some structural properties (Assumption 2 a)) allowing to construct linear paths that leave such boundaries. These paths can be realized as a sequence of jumps $\mathcal{E}$ whose rates do not decay too fast (Assumption 2 b)-c)), as to avoid for the realization of such path to have an infinite large-deviation cost. Denoting throughout by $\mathcal{B}_\varrho(x)$ the euclidean ball of radius ϱ in $\mathbb{R}^d$ and by $|A|$ the Lebesgue measure of the set A, we summarize such assumptions below:

Assumption 2 (Escape)**.** *There exist constants $\varepsilon, \varepsilon', \varepsilon'' > 0$ such that for each $j \in \mathcal{I}$, the following holds:*

a) If $j \in \mathcal{I}^{bd}$ there is a $w_j \in \mathbb{R}^d$ with $\|w_j\| = 1$ and $\kappa_j \in (0,1)$ such that whenever $x \in \mathcal{A}_j$ and $\inf_{y\in\partial\mathcal{S}} \|x - y\| < \varepsilon'$ and $t \in (0, \varepsilon)$

i) $t \mapsto \inf_{y\in\partial\mathcal{S}} \|x + t\, w_j - y\|$ is increasing, and

ii) $\mathcal{B}_{t\kappa_j}(x + t\, w_j) \cap \partial\mathcal{S} = \varnothing$.

We write $\kappa_- = \min_{j\in\mathcal{I}^{bd}}\{\kappa_j\}$. If $j \in \mathcal{I}\setminus\mathcal{I}^{bd}$ we choose $w_j = 0$.

b) There exists a finite sequence $\mathcal{E}_j := (r_1, \ldots, r_{n_j})$ of jumps in $\mathcal{R}$ with

$$\limsup_{v\to\infty} \frac{1}{v}\left|\log \Lambda_{r_k}^v\left(x_0^v + v^{-1}\sum_{i=1}^{k-1}\gamma^{r_i}\right)\right| = 0, \qquad k = 1, \ldots n_j$$

and $\sum_{i=1}^n \gamma^{r_i} = \alpha_j w_j$ for some $\alpha_j > 0$.

c) *Defining* $Z_0^v := \{x_0^v + v^{-1}\sum_{i=1}^k \gamma^{r_i} \ : \ k \in (0,\dots,n_j-1)\}$*, for all* $r \in \mathcal{E}_j$ *and* $T > 0$

$$\lim_{\varrho\to 0} \sup_{x\in\mathcal{A}_j\cup\bigcup_v Z_0^v} \int_0^\varrho |\log \lambda_r(x+sw_j)| \,\mathrm{d}s = 0 \,.$$

d) *Let* $\mathcal{W}_{j,\kappa''} := \{w_j + y \ : \ \|y\| < \kappa''\}$*. There exists* $\kappa'' < \kappa_-/3$ *such that for any* $x \in \mathcal{A}_j \cup \bigcup_v Z_0^v$ *we have that for all* $r \in \mathcal{R}$ *with* $\lambda_r(x) < \varepsilon''$ *the rates* $\lambda_r(\,\cdot\,)$ *are nondecreasing along paths* $t \mapsto x + tw$ *for any* $w \in \mathcal{W}_{j,\kappa''}$*, for* $t \in (0,\varepsilon)$*.*

It is readily verified Assumption 1 and 2 are satisfied by mass action kinetics rates on a convex domain [ADE18a].

Remark 2.5. *While Assumption 2 c) is natural in terms of our proof, we note that it is automatically satisfied whenever there exists* $\alpha \in [0,1)$ *such that*

$$\lim_{\varrho\to 0} \varrho^\alpha \left(\inf_{x\in\mathcal{A}_j\cup\bigcup_v Z_0^v:\, d(x,\partial\mathcal{A}_j)>\varrho} \log \lambda_r(x) \right) \to 0 \quad \text{for all } r \in \mathcal{E}_j \,, \tag{2.3}$$

as we mentioned in (1.3)*, where* $d(A,B) := \inf_{x\in A, y\in B} \|x-y\|$*. In particular, the above decay condition implies the results of [SW05] and [PSK16]. These papers make the stronger assumptions that (2.3) holds with* $\alpha = 0$ *and* $\alpha \in [0,1/2)$ *respectively.*

2.2 The large-deviation principle

For a parameter $T > 0$ fixed throughout the paper, we denote by $D_u(0,T;\mathbb{R}^d)$ (resp. $D_s(0,T;\mathbb{R}^d)$) the space of càdlàg functions with values in $\mathbb{R}^d$ endowed with the topology of uniform convergence (resp. Skorohod topology). Furthermore we define $\mathrm{B}_{[0,T]}(\varrho,z)$ to be the ball of radius ϱ in $D_u(0,T;\mathbb{R}^d)$. Finally, for $z\colon [0,T] \mapsto \mathbb{R}^d$ in the set $\mathcal{AC}(0,T;\mathbb{R}^d)$ of absolutely continuous functions, we denote by $\dot z$ its time derivative and we will say that $z \in \mathcal{AC}(0,T;\mathcal{S})$ whenever $z(t) \in \mathcal{S}$ a.e. $t \in [0,T]$.

To define the standard rate function for the LDP of Markov jump processes in the small noise limit [SW95] we introduce the action

$$I_{[0,T]}(z) := \begin{cases} \int_0^T \inf_{\left\{\mu\in\mathbb{R}^\mathcal{R}:\, \sum_{r\in\mathcal{R}} \mu_r\gamma^r = \dot z\right\}} \mathcal{H}(\mu|\lambda(z(t)))\,\mathrm{d}t, & \text{if } z \in \mathcal{AC}(0,T;\mathcal{S})\,, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.4}$$

$$\mathcal{H}(\mu|\lambda) := \sum_{r\in\mathcal{R}} \lambda_r - \mu_r + \mu_r \log\frac{\mu_r}{\lambda_r}\,, \tag{2.5}$$

where $(\lambda(x))_r := \lambda_r(x)$. We can now state the main result of this paper:

Theorem 1. *Consider the sequence of Markov jump processes* $\{X^v\}_{v\in\mathbb{N}}$*, fixing a sequence of deterministic initial conditions* $\{x_0^v\}_{v\in\mathbb{N}}^\infty$ *with* $x_0^v \in (v^{-1}\mathbb{Z})^d$ *such that* $X^v(0) = x_0^v \to x_0 \in \mathbb{R}^d$*. Furthermore, let Assumption 1 and 2 hold. Then the sequence* $\{X^v\}_{v\in\mathbb{N}}$ *satisfies a* LDP *in* $D_u(0,T;\mathbb{R}^d)$ *(resp.* $D_s(0,T;\mathbb{R}^d)$*) with good rate function*

$$I^{x_0}_{[0,T]}(z) := \begin{cases} I_{[0,T]}(z) & \text{if } z(0) = x_0 \\ +\infty & \text{otherwise,} \end{cases} \tag{2.6}$$

that is, $I^{x_0}_{[0,T]}$ has compact sublevel sets $D_u(0,T;\mathbb{R}^d)$ (resp. $D_s(0,T;\mathbb{R}^d)$), and for all measurable $\Gamma \subseteq D_s(0,T;\mathbb{R}^d)$,

$$\limsup_{v\to\infty} \frac{1}{v} \log \mathbb{P}_{x_0^v}\left[X^v \in \Gamma\right] \le - \inf_{x\in\bar{\Gamma}} I^{x_0}_{[0,T]}(x) \tag{2.7}$$

$$\liminf_{v\to\infty} \frac{1}{v} \log \mathbb{P}_{x_0^v}\left[X^v \in \Gamma\right] \ge - \inf_{x\in\Gamma^o} I^{x_0}_{[0,T]}(x), \tag{2.8}$$

where $\mathbb{P}_{x_0^v}\left[\,\cdot\,\right]$ denotes the conditional probability on $X_0^v = x_0^v$.

The compactness of $\mathcal{S}$ and Lipschitz continuity of the rates $\{\lambda_r\}$ directly implies that the rates are bounded and Lipschitz, which in turn implies that the process is exponentially tight. Therefore, the compactness of sublevel sets of $I^{x_0}_{[0,T]}$ comes for free, and the upper bound only needs to be proven for compact sets [DZ87, Lem. 1.2.18].

A few considerations are now in order.

Remark 2.6. *The boundedness of the rates and compactness of $\mathcal{S}$ can be relaxed. Indeed, exponential tightness can be obtained by other means: Either by Lipschitz continuity of the jump rates [FK06, SW95], or by stability estimates [ADE18a,* **?***]. Once exponential tightness is guaranteed, one can restrict the analysis to trajectories that do not leave a large enough compact [FK06, Theorem 4.4], effectively reducing the problem to the one with compact state space, which we discuss above.*

We further note that the seemingly restrictive assumption of *deterministic* initial condition also covers the case when such an initial condition is random. This can be done, given the probability conditioned on a fixed initial state from Theorem 1, by integrating with respect to the probability distribution $\nu^v \in \mathcal{M}((v^{-1}\mathbb{Z})^d)$ of the initial condition, provided that the measure ν^v satisfies some weak regularity and tightness assumptions. In this case, however, one must check that the conditions in Theorem 1 hold *uniformly* on a set of positive measure WRT ν^v. For a detailed discussion of this procedure when ν also satisfies a LDP at the same rate we refer to [Big04].

3 Proof of LDP lower bound

The general strategy adopted to prove the LDP lower bound result is mainly standard. Without loss of generality, we may assume that Γ is open, for any path $z \in \Gamma$ one can find a $\delta > 0$ such that $\mathrm{B}_{[0,T]}(\delta, z) \subset \Gamma$, so that $\mathbb{P}_x\left[X^v \in \Gamma\right] \ge \mathbb{P}_x\left[\mathrm{B}_{[0,T]}(\delta, z)\right]$ for $\delta > 0$ small enough. Hence, it is sufficient to prove that, for any path $z \in \Gamma$ the probability that the process X^v stays in a neighborhood $\mathrm{B}_{[0,T]}(\delta, z)$ for any $\delta > 0$ is approximately $\exp[-v I^{x_0}_{[0,T]}(z)]$. Applying such estimate to a sequence $\{z^{(n)}\}_{n=1}^\infty$ of paths converging to the minimizer of $I^{x_0}_{[0,T]}$ in Γ with small enough $\delta^{(n)}$ proves the desired result. This shows that for the lower bound (2.8) it is sufficient to prove the following.

Proposition 3.1. *Fix a path $z : [0,T] \to \mathcal{S}$ with a fixed initial condition $z(0) = x_0 \in \mathcal{S}$ such that $I^{x_0}_{[0,T]}(z) = K < \infty$. Then, for a sequence of initial conditions $x_0^v \in (v^{-1}\mathbb{Z})^d$ converging to x_0, under Assumption 1 and Assumption 2,*

$$\lim_{\delta\to 0} \liminf_{v\to\infty} \frac{1}{v} \log \mathbb{P}_{x_0^v}\left[X^v \in \mathrm{B}_{[0,T]}(\delta, z)\right] \ge -I^{x_0}_{[0,T]}(z)\,. \tag{3.1}$$

The remainder of this section concentrates on proving such estimate. Our approach to the proof of the above result mimicks the one from [SW05]. Throughout this section, we fix a path $z \in \mathcal{AC}([0,T],\mathcal{S})$ starting from $z(0) = x_0$ and we approximate z with another path $\mathfrak{z}_\delta$ obtained by perturbing z, shifting it uniformly away from the regions where the rates are degenerate by a quantity controlled by δ. We then proceed to prove on one hand that the probability of X^v approximately following $\mathfrak{z}_\delta$ is accurately described by the rate function $I^{x_0}_{[0,T]}(\mathfrak{z}_\delta)$, and on the other that the large-deviation cost of the process following the *shifted* path converges towards the one of the original path as $\delta \to 0$. The main difficulty to establish the former claim arises with the necessity of keeping the microscopic initial condition of the path fixed, and estimating the probability of the process reaching, in a small time interval, the origin of the shifted path $\mathfrak{z}$, which is *macroscopically* bounded away from the boundary. On the other hand, to establish the latter convergence property of the rate functional we have to guarantee sufficient regularity of the rate functional as some of the jump rates decrease to 0 with $\delta \to 0$. The remainder of the section is devoted to the realization of this program. In Section 3.1 we give the explicit construction of the path $\mathfrak{z}_\delta$ and detail its role in the proof of Proposition 3.1, in Section 3.2 we estimate the probability of the process reaching the origin of the shifted path from its fixed initial condition x_0, while in Section 3.3 we prove sufficient regularity of the rate functional $I^{x_0}_{[0,T]}$. The proof of Proposition 3.1 is finally concluded in Section 3.4.

3.1 Construction of the path $\mathfrak{z}_\delta$

We now construct a macroscopic path that perturbs the original path $z \in \mathcal{AC}(0,T;\mathcal{S})$ at a negligible cost and that can only intersect $\partial\mathcal{S}$ at its initial point. To do so we recall the covering $\{\mathcal{A}_i\}$ defined in Section 2.1, allowing us to identify, for each $\mathcal{A}_i$, directions w_i to move away from the boundary $\partial\mathcal{S}$. More specifically, this covering allows to partition the path z as it enters different regions $\mathcal{A}_j$ and to shift it in the corresponding direction w_j, thereby guaranteeing that the *shifted* path avoids $\partial\mathcal{S}$ as we detail below and as depicted in Fig. 1.

To construct $\mathfrak{z}_\delta$ we introduce the sequence of times $\{\tau_k\}_k$ so that $\{z(t) \;:\; t \in [\tau_k, \tau_{k+1}]\} \subset \mathcal{A}_{i_k}$ for all k for a corresponding sequence $\{i_k\}_k$ of indices in $\mathcal{I}$. Then, for fixed $x_0 \in \mathcal{S}$, we consider $i_0 \in \mathcal{I}$ such that $x_0 \in \mathcal{A}_{i_0}$, so that $z(t) \in \mathcal{A}_{i_0}$ for all $t \in [0,\tau_1]$. In this interval we define the shifted path

$$\mathfrak{z}_\delta(t) := \begin{cases} x_0 + t w_j & \text{for } t \in [0,t_\delta] \\ z(t-t_\delta) - x_0 + t_\delta w_j & \text{for } t \in (t_\delta, \tau_1 + t_\delta] \end{cases} \tag{3.2}$$

for $t_\delta = \frac{1}{6}\min(\delta\xi, \omega_z^{-1}(\delta))$ where ω_z denotes the modulus of continuity of the path z and $\xi > 0$ is defined below (see Lemma 3.2). We then continue defining the path $\mathfrak{z}_\delta$ by shifting the original path z infinitesimally on every interval $[\tau_k, \tau_{k+1}]$, sequentially moving it away from $\partial\mathcal{A}_{i_k}$ with the corresponding w_{i_k}. More precisely, setting the length of the k-th shift time for the perturbed path as $\beta^k t_\delta$ for $\beta > 1$ to be chosen later (see Lemma 3.2) and denoting by $\Delta_k^\beta := t_\delta \sum_{\ell=0}^{k-1} \beta^\ell$ the cumulative shift time up to transition k we define the i-th shift as

$$\mathfrak{z}_\delta(s) := \begin{cases} \mathfrak{z}_\delta(t) + (s - \tau_i - \Delta_i^\beta) w_i & \text{for } s \in (\tau_i + \Delta_i^\beta, \tau_i + \Delta_{i+1}^\beta] \\ \mathfrak{z}_\delta(\tau_i + \Delta_i^\beta) + \beta^k t_\delta w_i + z(s - \Delta_{i+1}^\beta) - z(\tau_i) & \text{for } s \in (\tau_i + \Delta_{i+1}^\beta, \tau_{i+1} + \Delta_{i+1}^\beta]. \end{cases} \tag{3.3}$$

We now establish some structural properties of the newly constructed path around the original z, which we recall is fixed throughout this section. This lemma extends [SW05, Lemma 3.4].

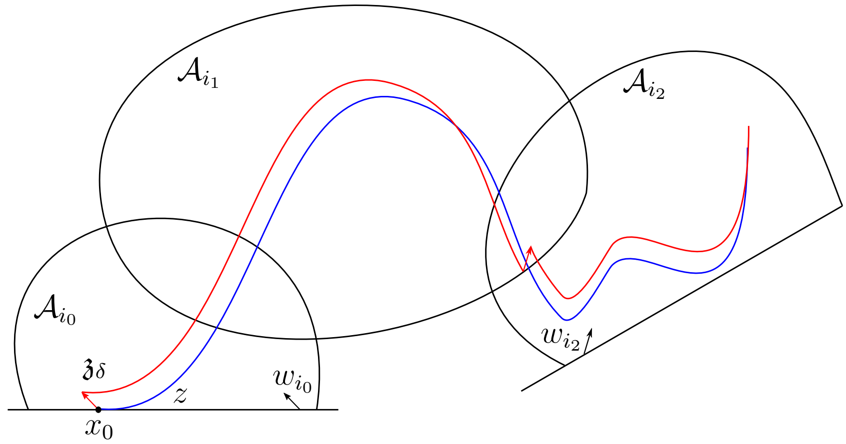

Figure 1: Schematic representation of shifted path.

Lemma 3.2. *Let Assumptions 1 and 2 hold and set* $\beta := 3/\kappa''$ *recalling that* $\kappa'' < \kappa_- = \min_{j\in\mathcal{I}} \kappa_j$ *from Assumption 2 a). Then, for any* $K > 0$ *there is a* $J > 0$ *such that if* $I_{[0,T]}(z) \leq K$*, there are* $0 = \tau_0 < \tau_1 < \cdots < \tau_J = T$ *and* $\{i_k\}$ *with* $z(t) \in \mathcal{A}_{i_k}$ *for* $\tau_{k-1} \leq t \leq \tau_k$*. Furthermore, setting* $\xi := \min(1, (\kappa''/3)^{J+1}/3, \varepsilon)$ *there exists* $\delta_z > 0$ *such that for all* $\delta < \delta_z$ *the path* $\mathfrak{z}_\delta$ *from (3.2) and* (3.3) *satisfies* $\sup_{[0,T]} \|z - \mathfrak{z}_\delta\| < 2\delta/3$*.*
Finally, the path $\mathfrak{z}_\delta$ *satisfies* $\bigcup_{t\in[t_\delta,T]} \mathcal{B}_{\kappa_- t_\delta}(\mathfrak{z}_\delta(t)) \cap \partial\mathcal{S} = \varnothing$ *and for every* $i \in (1,\ldots,J)$ *and* $a \in \mathbb{R}^d \cap \mathrm{span}_{r\in\mathcal{R}}(\gamma^r)$ *with* $\|a\| < t_\delta\kappa''/2$ *there exists* $w \in \mathcal{W}_{i_k,\kappa''}$ *such that* $\mathfrak{z}_\delta(\tau_k + \Delta^\beta_{k+1}) + a = z(\tau_k) + \beta^k t_\delta w$*.*

We defer the proof of this lemma to the end of the section and proceed to present the central estimate allowing us to bound the probability in (3.1) from below — in the sense of large deviations. To do so, defining throghout $\beta := 3/\kappa''$ and $\xi := \min(1, (\kappa''/3)^{J+1}/3, \varepsilon)$ so that Lemma 3.2 holds, by triangle inequality it is sufficient to consider the event $\sup_{t\in[0,T]} \|\mathfrak{z}_\delta(t) - X^v(t)\| < \delta/3$. Furthermore, $\mathfrak{z}_\delta \in \mathcal{S}$ and for any $\delta''/2 \leq \delta' \leq \delta/3$ we can further bound the event of interest from below as follows:

$$\begin{aligned}
\mathbb{P}_{x_0^v}\Big[X^v \in \mathrm{B}_{[0,T]}(\delta, z)\Big] &\geq \mathbb{P}_{x_0^v}\left[X^v \in \mathrm{B}_{[0,T]}(\delta/3, \mathfrak{z}_\delta)\right] \\
&\geq \mathbb{P}_{x_0^v}\left[\{X^v \in \mathrm{B}_{[0,t_\delta]}(\delta/3, \mathfrak{z}_\delta)\} \cap \{X^v(t_\delta) \in \mathcal{B}_{\delta''/2}(x_0 + t_\delta w_j)\} \cap \{X^v \in \mathrm{B}_{[t_\delta,T]}(\delta', \mathfrak{z}_\delta)\}\right] \\
&\geq \mathbb{P}_{x_0^v}\left[\{X^v \in \mathrm{B}_{[0,t_\delta]}(\delta/3, \mathfrak{z}_\delta)\} \cap \{X^v(t_\delta) \in \mathcal{B}_{\delta''/2}(x_0 + t_\delta w_j)\}\right] \\
&\qquad \times \inf_{y\in\mathcal{B}_{\delta''/2}(x_0+t_\delta w_j)} \mathbb{P}\left[X^v \in \mathrm{B}_{[t_\delta,T]}(\delta', \mathfrak{z}_\delta) \mid X^v(t_\delta) = y\right] ,
\end{aligned} \tag{3.4}$$

where in the last inequality we have used the Markov property. In the remainder of the paper we set

$$\delta' := \kappa_- t_\delta/3 \qquad \text{and} \qquad \delta'' := t_\delta\kappa'' < \delta'$$

where recalling that $\kappa_- = \min_{j\in\mathcal{I}^{bd}}\{\kappa_j\} < 1$ and that $\xi < 1$ we must have $\delta' \leq \delta/3$. We note that this choice is compatible with the definition of κ'' from Assumption 2.

Remark 3.3. *We pause briefly to motivate our choice of* δ' *and* δ''*: These small parameters are chosen in such a way as to guarantee that the event in the second term in the last line of (3.4) only contains paths that are uniformly bounded away from the boundary* $\partial\mathcal{S}$*, as captured by Lemma 3.2 and depicted in Fig. 2.*

The desired result is obtained by showing that

$$\lim_{\delta\to 0}\liminf_{v\to\infty}\frac{1}{v}\log\mathbb{P}_{x_0^v}\left[X^v\in \mathrm{B}_{[0,t_\delta]}(\delta/3,\mathfrak{z}_\delta)\cap\{X^v(t_\delta)\in\mathcal{B}_{\delta''/2}(x_0+w_jt_\delta)\}\right]=0 \qquad \text{and} \tag{3.5}$$

$$\lim_{\delta\to 0}\liminf_{v\to\infty}\frac{1}{v}\log\inf_{y\in\mathcal{B}_{\delta''/2}(x_0+t_\delta w)}\mathbb{P}\left[X^v\in \mathrm{B}_{[t_\delta,T]}(\delta',\mathfrak{z}_\delta)\mid X^v(t_\delta)=y\right]\geq -I^{x_0}_{[0,T]}(z). \tag{3.6}$$

The term in (3.5) is bounded from below in Section 3.2, while in Section 3.3 and Section 3.4 we formulate and combine the estimates in different $\mathcal{A}_j$ to bound (3.6), thereby proving the desired LDP lower bound.

3.2 Jump bounds

We now proceed to bound from below the first term on the last line of (3.4). To do this we consider a convenient subset of outcomes obtained by fixing a precise sequence of jumps (but not the *times* of the jumps) that the process undergoes in the interval $(0,t_\delta)$. To define such an event, we recall the definition of the sequence $\mathcal{E}_j$ of n_j jumps leading away from $\partial\mathcal{A}_j$ and we denote the event of repeating the sequence of jumps in $\mathcal{E}_j$ n times by

$$\Xi_j(n,v):=\bigcap_{m=0}^{n-1}\bigcap_{i=1}^{n_j}\left\{X^v(\sigma_{mn_j+i})-X^v(\sigma_{mn_j+i}-)=v^{-1}\gamma^{r_i}\right\}, \tag{3.7}$$

where, for all $k\in\mathbb{N}$, σ_k is the time of the k-th jump of the Markov process X^v. Furthermore, we note that by our choice of $\xi<1/6$ we must have $\{x_0+tw_j\ :\ t\in(0,t_\delta)\}\subset\bigcap_{t\in(0,t_\delta)}\mathcal{B}_{\delta/3}(x_0+w_jt)$, as depicted in Fig. 2. Thus for $n_+^v:=\lfloor\frac{v}{\alpha_j}t_\delta\rfloor$ and $n_-^v:=\lceil\frac{v}{\alpha_j}(t_\delta-\delta')\rceil$ we have for all v large enough that $n_-^v<n_+^v$,

$$\{X^v\in\mathrm{B}_{[0,t_\delta]}(\delta/3,\mathfrak{z}_\delta)\}\supseteq\Xi_j(n_+^v,v)\cap\{\sigma_{n_+^vn_j}>t_\delta\}, \tag{3.8}$$

and also

$$\{X^v(t_\delta)\in\mathcal{B}_{\delta'}(x_0+w_jt_\delta)\}\supseteq\Xi_j(n_+^v,v)\cap\{\sigma_{n_+^vn_j}>t_\delta\}\cap\{\sigma_{n_-^vn_j}\leq t_\delta\}.$$

Note that, as v and n increase the paths in $\Xi_j(n,v)$ have ranges concentrating on a straight line segment in $\mathbb{R}^d$ However, there is no information about the speed at which they move along this line segment. This degree of freedom will be sufficient to establish the LDP lower bound, as we shall see next. In preparation for the next result observe by Assumption 1a) that $\limsup_{v\to\infty}\sup\left\{\sum_{r\in\mathcal{R}}\Lambda_r^v(x)\colon x\in\mathcal{B}_{2t_\delta}(x_0)\right\}<\infty$ and define $\bar{t}_\delta(\alpha,\varepsilon''):=\min_j\{\alpha_j/n_j,\varepsilon''/\max_{r\in\mathcal{R}}\mathrm{Lip}(\lambda_r)\}$.

Lemma 3.4. *Suppose $x_0\in\mathcal{A}_j$ and let Assumption 1 and 2 hold and that δ is small enough that $t_\delta<\bar{t}_\delta(\alpha,\varepsilon'')$, then for $\bar{\lambda}>\max\{1,\limsup_{v\to\infty}\sup\left\{\sum_{r\in\mathcal{R}}\Lambda_r^v(x)\colon x\in\mathcal{B}_{2t_\delta}(x_0)\right\}\}$ we have*

$$\begin{aligned}\liminf_{v\to\infty}\frac{1}{v}\log\mathbb{P}&\left[X^v(t_\delta)\in\mathcal{B}_{\delta''/2}(x_0+t_\delta w_j)\,,\,\Xi_j(n^v,v)\,,\,\sigma_{n_+^vn_j}>t_\delta\middle|X^v(0)=x_0^v\right]\\&\geq -t_\delta\left(\frac{n_j}{\alpha_j}\log\left(\frac{n_j}{\alpha_j\bar{\lambda}}\right)-\frac{n_j}{\alpha_j}+\bar{\lambda}\right)+\sum_{i=1}^{n_j}\int_0^{t_\delta/\alpha_j}\log\left(\lambda_{r_i}\left(x_0+s\alpha_jw_j\right)\right)\mathrm{d}s+t_\delta\frac{n_j}{\alpha_j}\log\aleph\,.\end{aligned}$$

To prove the above result, defining throughout $\tilde{\gamma}^{(i)}:=\sum_{k=1}^{i-1}\gamma^{r_k}$ we introduce the following lemma relating the Riemann sum of $\log\lambda_r$ along the escape sequence defining Ξ to the corresponding integral.

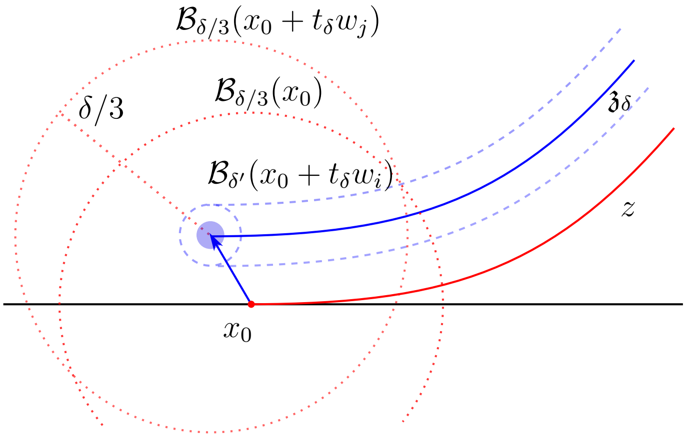

Figure 2: Schematic representation of the desired effect for the choice of parameters $\delta, \delta', \delta''$ summarized in Lemma 3.2. For a fixed path z and $\delta > 0$ by our choice of $\xi > 0$ and consequently t_δ we have that a neighborhood of the path $x_0 + w_{i_0} t$ (blue arrow) is contained in $\bigcap_{t \in (0,t_\delta)} \mathcal{B}_{\delta/3}(x_0 + t w_{i_0})$ (the intersection of the two dotted balls). By our choice of $\delta'(\xi, \kappa_-) > 0$, we find that a δ'-neighborhood (dashed blue region) of the path $\mathfrak{z}_\delta$ (blue line) on $[t_\delta, T)$ never intersects ∂S while $\mathcal{B}_{\delta'}(x_0 + w_{i_0} t_\delta) \subset \bigcap_{t \in (0,t_\delta)} \mathcal{B}_\delta(z(t))$. The shaded blue region represents $\mathcal{B}_{\delta''/2}(x_0 + t_\delta w_j)$.

Lemma 3.5. *Suppose $x_0 \in \mathcal{A}_j$, let Assumption 1 and 2 hold, then for all δ such that $t_\delta < \bar{t}_\delta(\alpha, \varepsilon'')$ and $r_i \in \mathcal{E}_j$ we have*

$$\liminf_{v \to \infty} \frac{1}{v} \sum_{m=0}^{n_+^v} \log \left(\lambda_{r_i} \left(x_0^v + \frac{m}{v} \alpha_j w_j + v^{-1} \tilde{\gamma}^{(i)} \right) \right) \geq \int_0^{t_\delta/\alpha_j} \log \left(\lambda_{r_i} \left(x_0 + s \alpha_j w_j \right) \right) \mathrm{d}s \,. \tag{3.9}$$

Proof of Lemma 3.4. When $x_0 \notin \partial \mathcal{A}_j$ all the rates are strictly positive by definition, so the result follows by standard large-deviation estimates [SW95]. Therefore, for the rest of the proof we assume that $x_0 \in \partial \mathcal{A}_j$. Denoting by $\mathbb{1}\{x \in A\}$ the indicator function on the set A, we introduce a jump $r^\star$ with

$$\Lambda_{r^\star}^v(x) := \left(\bar{\lambda} - \sum_{r \in \mathcal{R}} \Lambda_r^v(x) \right) \mathbb{1}\{x \in \mathcal{B}_{2t_\delta}(x_0)\} \qquad \text{and} \qquad \gamma^{r^\star} = 0 \,,$$

and expand the set of jumps $\mathcal{R}^\star := \mathcal{R} \cup \{r^\star\}$. We then define a new family of processes $\bar{X}^v$ on the extended set of jumps $\mathcal{R}^\star$ and corresponding jump rates. By independence of the jump processes we trivially couple the underlying Poisson processes for jumps in $\mathcal{R}$ to the ones of the process X^v, so that $\bar{X}^v(t) = X^v(t)$ a.s. and proceed to establish the desired result for $\bar{X}^v$. In the rest of the proof, by abuse of notation we will denote by $\bar{\Xi}_j$ the set defined in (3.7) for $\bar{X}^v$ instead of X^v. We then have

$$\begin{aligned} &\mathbb{P}\left[\bar{X}^v(t_\delta) \in \mathcal{B}_{\delta''/2}(x_0 + t_\delta w_j), \sigma_{n_+^v n_j} > t_\delta, \bar{\Xi}_j(n_+^v, v) \middle| \bar{X}^v(0) = x_0^v \right] \\ &\qquad \geq \mathbb{P}\left[\sigma_{n_-^v n_j} \leq t_\delta < \sigma_{n_+^v n_j} \middle| \bar{\Xi}_j(n_+^v, v), \bar{X}^v(0) = x_0^v \right] \times \mathbb{P}\left[\bar{\Xi}_j(n_+^v, v) \middle| \bar{X}^v(0) = x_0^v \right] . \end{aligned}$$

On the event $\bar{\Xi}_j(n_+^v, v) \cap \left\{ \sigma_{n_+^v n_j} \geq t \right\}$ one has, for v large enough,

$$\sup_{s<t} \left\| \bar{X}^v(s) - x_0 \right\| \leq \|x_0 - x_0^v\| + t_\delta \|w_j\| + v^{-1} \max_i \left\| \tilde{\gamma}^{(i)} \right\| < 2t_\delta \,,$$

and therefore $\sum_{r\in\mathcal{R}^*}\Lambda_r^v(\bar{X}^v(s))=\bar{\lambda}$ for all $s<t$.

Now, recalling that $\alpha_j w_j := \sum_{k=1}^{n_j}\gamma^{r_k}$ with $\|w_j\|=1$, by the conditional independence of the jumps

$$\mathbb{P}\left[\bar{\Xi}_j(n_+^v,v)\middle|\bar{X}^v(0)=x_0^v\right] = \prod_{m=0}^{n_+^v-1}\prod_{i=1}^{n_j}\frac{\Lambda_{r_i}^v\left(x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)}{\bar{\lambda}}\,. \tag{3.10}$$

and thus

$$\begin{aligned}\frac{1}{v}\log\mathbb{P}\left[\bar{\Xi}_j(n_+^v,v)\middle|\bar{X}^v(0)=x_0^v\right] \geq \sum_{i=1}^{n_j}\frac{1}{v}\sum_{m=0}^{n_+^v-1}\log\left(\frac{\Lambda_{r_i}^v\left(x_0^v+mv^{-1}\beta_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)}{\lambda_{r_i}\left(x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)}\right)\\ +\sum_{i=1}^{n_j}\frac{1}{v}\sum_{m=0}^{n_+^v-1}\log\left(\frac{\lambda_{r_i}\left(x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)}{\bar{\lambda}}\right)\,.\end{aligned} \tag{3.11}$$

Now, using Lemma 3.5 and recalling the definition of $\aleph>0$ from Assumption 1 we have for (3.11)

$$\begin{aligned}&\liminf_{v\to\infty}\frac{1}{v}\log\mathbb{P}\left[\bar{\Xi}_j(n^v,v)\middle|\bar{X}^v(0)=x_0^v\right]\\ &\qquad=\sum_{i=1}^{n_j}\int_0^{t_\delta/\alpha_j}\log\Big(\lambda_{r_{\ell_i}}\left(x_0+s\alpha_j w_j\right)\Big)\,\mathrm{d}s-n_j t_\delta\log\left(\bar{\lambda}/\aleph\right)/\alpha_j.\end{aligned} \tag{3.12}$$

Also since the waiting times between jumps are independent of which type of jump actually occurs

$$\mathbb{P}\left[\bar{X}^v(t_\delta)\in\mathcal{B}_{\delta''/2}(x_0+t_\delta w_j),\sigma_{n_+^v n_j}>t_\delta\middle|\bar{\Xi}_j(n^v,v),\bar{X}^v(0)=x_0^v\right]=\mathbb{P}\left[n_-^v n_j\leq Y^v<n_+^v n_j\right],$$

where Y^v is Poisson distributed with mean $t_\delta v\bar{\lambda}$. By definition, $\bar{\lambda}\geq 1$ so that it is in particular greater than $n_+^v n_j/v$ for our choice of t_δ, and we have

$$\lim_{v\to\infty}\frac{1}{v}\log\mathbb{P}\left[n_-^v n_j\leq Y^v<n_+^v n_j\right]=-t_\delta\left(\frac{n_j}{\alpha_j}\log\left(\frac{n_j}{\alpha_j\bar{\lambda}}\right)-\frac{n_j}{\alpha_j}+\bar{\lambda}\right) \tag{3.13}$$

The result now follows by combining (3.13) and (3.12). ☐

Proof of Lemma 3.5. We note that λ_r is Lipschitz so whenever $\lambda_r(x_0)>0$ then $\lambda_r(x+sw_j)$ is uniformly bounded away from 0 on a sufficiently small time interval and the result follows immediately by a dominated convergence argument. We therefore assume throughout that $\lambda_r(x_0)=0$. Introducing

$$\tilde{n}^v := \inf\left\{m\in\mathbb{N}\ :\ \frac{x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}-x_0}{\|x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}-x_0\|}\in\mathcal{W}_{j,\kappa''}\right\}, \tag{3.14}$$

we split the sum in the statement of the Lemma into the terms $m=0$, $m\in(0,\ldots,\tilde{n}^v)$ and $m\in(\tilde{n}^v+1,\ldots,n_+^v)$ and proceed to bound their constribution separately. The term $m=0$ is automatically bounded by Assumption 2b). For the second term we observe since $\log\left(\lambda_{r_i}\left(x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)\right)$ is increasing in m by Assumption 2d) that

$$\begin{aligned}&\liminf_{v\to\infty}\frac{1}{v}\sum_{m=1}^{\tilde{n}^v}\log\left(\lambda_{r_i}\left(x_0^v+mv^{-1}\alpha_j w_j+v^{-1}\tilde{\gamma}^{(i)}\right)\right)\\ &\qquad\geq\liminf_{v\to\infty}\frac{1}{v}\int_0^{\tilde{n}^v/v}\log\left(\lambda_{r_i}\left(x_0^v+v^{-1}\tilde{\gamma}^{(i)}+t\alpha_j w_j\right)\right)\mathrm{d}t=0\,,\end{aligned}$$

where the final equality arises since $\lim_{v\to\infty} x_0^v = x_0$ implies $\lim_{v\to\infty} v^{-1}\tilde{n}^v = 0$, and we have the integral estimate from Assumption 2c), which is uniform in the starting points $x_0^v + v^{-1}\tilde{\gamma}^{(i)} \in Z_0^v$.

Similarly for the terms $m \in (\tilde{n}^v + 1, \dots, n_+^v)$, defining $m' := m - \tilde{n}^v \geq 0$ we can write

$$x_0^v + mv^{-1}\alpha_j w_j + v^{-1}\tilde{\gamma}^{(i)} = x_0 + m'v^{-1}\alpha_j w_j + \left(x_0^v + \tilde{n}^v v^{-1}\alpha_j w_j + v^{-1}\tilde{\gamma}^{(i)} - x_0\right) , \quad (3.15)$$

and note that by (3.14) the vector in brackets, once renormalized, is in $\mathcal{W}_{j,\kappa''}$.
We also have $\lim_{v\to\infty} \left\|x_0^v + \tilde{n}^v v^{-1}\alpha_j w_j + v^{-1}\tilde{\gamma}^{(i)} - x_0\right\| = 0$ so provided that δ is small enough that $x_0 + m'v^{-1}\alpha_j w_j \in \mathcal{A}_j$ and $\lambda_{r_i}(x_0 + m'v^{-1}\alpha_j w_j) < \varepsilon''$ we may apply Assumption 2d) to see that for each $m' \geq 0$

$$\lambda_{r_i}\left(x_0^v + mv^{-1}\alpha_j w_j + v^{-1}\tilde{\gamma}^{(i)}\right) \geq \lambda_{r_i}\left(x_0 + m'v^{-1}\alpha_j w_j\right).$$

A second application of Assumption 2d) implies

$$\liminf_{v\to\infty} \frac{1}{v} \sum_{m'=1}^{n_+^v - \tilde{n}^v} \log\left(\lambda_{r_i}\left(x_0 + mv^{-1}\alpha_j w_j\right)\right) \geq \liminf_{v\to\infty} \int_0^{(n_+^v - \tilde{n}^v)/v} \log\left(\lambda_{r_i}\left(x_0 + t\alpha_j w_j\right)\right) \mathrm{d}t.$$

□

3.3 Approximation of the rate functional

After the process has left the boundary $\partial\mathcal{S}$, we estimate the cost of a given path z by approximating z with another path, that is uniformly bounded away from $\partial\mathcal{S}$. This implies that a standard LDP holds for such shifted path, and this can then be used to bound the rate function of the original z. We start by proving an adaptation of [SW05, Lemma 4.1] to the present setting, recalling that ω_z is the modulus of continuity of z.

Lemma 3.6. *Under Assumption 2, for every $x \in \mathcal{A}_j$ for $j \in \mathcal{I}$ recalling that $t_\delta := \min(\omega_z^{-1}(\delta), \xi\delta)$ the cost of the path $z_\delta(t) = x + tw_j$ satisfies*

$$\lim_{\delta\to 0} I_{[0,t_\delta]}(z_\delta) = 0\,.$$

Proof. The thesis follows immediately by Assumption 2, resulting in the integrability of the rate functional $I_{[0,t_\delta]}$ along the chosen trajectories. □

We define throughout

$$\ell(x,y) := \sup_{\vartheta\in\mathbb{R}^d} \vartheta\cdot y - \sum_{r\in\mathcal{R}} \lambda^r(x)\left(\exp\left(\vartheta\cdot\gamma_r\right) - 1\right). \quad (3.16)$$

and recall that, by convex duality, for any $x, y \in \mathbb{R}^d$ we have $\ell(x,y) = \inf_{\{\mu\in\mathbb{R}_{\geq 0}^{|\mathcal{R}|}\,:\,\sum_{r\in\mathcal{R}}\mu_r\gamma^r = y\}} \mathcal{H}(\mu|\lambda(x))$ for $\mathcal{H}(\mu|\lambda)$ defined in (2.5), as proven *e.g.*, in [SW95]. Consequently we can express $I_{[0,T]}(z) = \int_0^T \ell(z(s), z'(s))\,\mathrm{d}s$. This allows to prove the following adaptation of [SW05, Lemma 5.1].

Lemma 3.7. *Let Assumptions 1 and 2 hold. Fix $i \in \mathcal{I}$, $\tau > 0$ and let the path z take values in $\mathcal{A}_i$ for $t \in [0,\tau]$ and satisfy $I_{[0,\tau]}(z) < K$ for $K < \infty$. Then for any $w \in \mathcal{W}_{i,\kappa''}$, $C_\beta > 0$ the shifted path $z_\delta(\cdot) = z(\cdot) + C_\beta t_\delta w$ satisfies*

$$\limsup_{\delta\to 0} I_{[0,\tau]}(z_\delta) \leq I_{[0,\tau]}(z)$$

Proof. We define $\ell_1(t) = \ell(z(t), z'(t))$ and $\ell_2(t) = \ell(z_\delta(t), z'_\delta(t))$ and denote by $(\mu^*_r(t))_{r\in\mathcal{R}}$ the optimizing set of jumps in (2.4) for the path z. This minimizer exists, because the sublevel sets for $\mu \mapsto \mathcal{H}(\mu|\lambda)$ for $\mathcal{H}$ from (2.5) are compact. Then we have that $\ell_2(t) \le \mathcal{H}(\mu^*(t)|\lambda(z_\delta(t)))$. On the other hand, by continuity of the asymptotic rates λ_r there exists a function $K_\lambda(\delta)$ with $\lim_{\delta\to 0} K_\lambda(\delta) = 0$ for which we have $|\lambda_r(z(t)) - \lambda_r(z_\delta(t))| < K_\lambda(\delta)$ for all $r \in \mathcal{R}$, so that

$$\ell_2(t) - \ell_1(t) \le \sum_{r\in\mathcal{R}} \lambda_r(z_\delta(t)) - \lambda_r(z(t)) + \mu^*_r(t) \log \frac{\lambda_r(z(t))}{\lambda_r(z_\delta(t))} \le |\mathcal{R}| K_\lambda(\delta) + \sum_{r\in\mathcal{R}} \mu^*_r(t) \log \frac{\lambda_r(z(t))}{\lambda_r(z_\delta(t))} .$$

We now bound the second term on the RHS from above depending on whether $\lambda_r(z(t)) > \varepsilon''$ from Assumption 2. If $\lambda_r(z(t)) \le \varepsilon''$, by the assumed increasing property of $\lambda_r(\cdot)$ along $x + sw$, we have $\log \lambda_r(z(t))/\lambda_r(z_\delta(t)) \le 0$. On the other hand, if $\lambda_r(z(t)) > \varepsilon''$ then for δ small enough

$$\log \frac{\lambda_r(z(t))}{\lambda_r(z_\delta(t))} \le \log \frac{\lambda_r(z(t))}{\lambda_r(z(t)) - K_\lambda(\delta)} \le \log \frac{\varepsilon''}{\varepsilon'' - K_\lambda(\delta)} \le \frac{2K_\lambda(\delta)}{\varepsilon''} ,$$

where in the last inequality we used $\log(1-x)^{-1} < 2x$ for x small enough. It remains to show that the contribution of the term $\mu^*_r(t)$ is bounded from above on the paths of interest. This result is obtained in the proof of [SW05, Lemma 5.1] by the convexity and asymptotic growth of the Lagrangian $\ell(x, y)$ in its second argument, proven in [SW05, Lemma 5.1], [SW95, Lemma 5.17] leveraging only the boundedness of the rates λ_r. Following the same argument we bound $\mu^*_r(t) \le C_0(1 + \ell_1(t))$ and we finally obtain

$$\ell_2(t) \le \ell_1(t) + K_\lambda(\delta)(C_1 + C_2\ell_1(t))$$

for sufficiently large, positive constants C_0, C_1, C_2. By the assumed boundedness of $I_{[0,\tau]}(z)$, this gives the desired result by integration. □

We now combine the above estimates, established in each region $\mathcal{A}_i$ separately, to obtain convergence of the rate functional $I_{[0,T]}(\mathfrak{z}_\delta)$ to $I_{[0,T]}(z)$ as $\delta \to 0$. While the idea of the proof is the same as in the original reference, we have to reproduce the process more closely, as in our case we cannot bound, in general, the rate function of the shifts linearly in δ as done in [SW05, Lemma 4.1] and we only have the limiting result Lemma 3.6. To bypass this issue, we leverage exponential tightness – discussed directly below the statement of Theorem 1 – to show that the number of transitions between different $\mathcal{A}_j$ done by the path of interest is bounded uniformly on sublevel sets of the rate functional. For any $a \in \mathbb{R}^d$, we extend the definition of $\mathfrak{z}_\delta$ on the interval $[0, T]$ as:

$$\tilde{\mathfrak{z}}^a_\delta(t) = \begin{cases} \mathfrak{z}_\delta(t_\delta) + a & \text{for } t \in [0, t_\delta) \\ \mathfrak{z}_\delta(t) + a & \text{for } t \in [t_\delta, T] \end{cases} .$$

Lemma 3.8. *Let Assumptions 1 and 2 hold, and let the path z satisfy $I_{[0,T]}(z) < \infty$. Then the path $\tilde{\mathfrak{z}}_\delta$ satisfies*

$$\limsup_{\delta\to 0} \sup_{a\in \mathrm{span}_{r\in\mathcal{R}}(\gamma^r)\,:\,\|a\|<\kappa'' t_\delta} I_{[0,T]}(\tilde{\mathfrak{z}}^a_\delta) \le I_{[0,T]}(z)$$

Proof. Recall the definition of times $\tau_1 \ldots, \tau_J$ from Lemma 3.2, separating $[0, T]$ in a finite number of intervals where the path z is contained in a set $\mathcal{A}_j$. We now express the rate function as the sum of the cost of the shifted path and the cost of the shifts: For $\Delta_i := t_\delta \sum_{k=0}^{i} (3/\kappa'')^k$ (reflecting the choice of β in Lemma 3.2) we write

$$I_{[0,T]}(\tilde{\mathfrak{z}}^a_\delta) = I_{[0,t_\delta)}(\tilde{\mathfrak{z}}^a_\delta) + \sum_{i=1}^{J} I_{[\tau_{i-1}+\Delta_i,\tau_i+\Delta_i]}(\tilde{\mathfrak{z}}^a_\delta) + \sum_{i=1}^{J} I_{[\tau_i+\Delta_i,\tau_i+\Delta_{i+1}]}(\tilde{\mathfrak{z}}^a_\delta)\,, \tag{3.17}$$

and proceed to bound the terms on the RHS separately. For the first term we can trivially choose the optimizing set of fluxes in (2.5) as $\mu^* = 0$, so that by boundedness of the rates $\lambda_r(x) < C$ we have $I_{[0,t_\delta)}(\tilde{\mathfrak{z}}^a_\delta) \le |\mathcal{R}| C t_\delta$, which vanishes with $\delta \to 0$.

We proceed to bound the summands in the second term. Recalling by Lemma 3.2 that for each time interval $[\tau_{i-1}+\Delta_i, \tau_i+\Delta_i]$ the trajectory of $\tilde{\mathfrak{z}}^a_\delta$ corresponds to the one of $z + wt_\delta$ for $w \in \mathcal{W}_{j_i,\kappa''}$ we see that for each such interval we can apply Lemma 3.7. Combining this result with the time-translation invariance of the rate functional we obtain that

$$\limsup_{\delta\to 0} \sum_{i=1}^{J} I_{[\tau_{i-1}+\Delta_i,\tau_i+\Delta_i]}(\tilde{\mathfrak{z}}^a_\delta) \le \sum_{i=1}^{J} I_{[\tau_{i-1},\tau_i]}(z)\,. \tag{3.18}$$

We then bound the third term of (3.17) by Lemma 3.6, recalling that by Lemma 3.2 the path $\tilde{\mathfrak{z}}^a_\delta$ is in $\mathcal{S}$. We start by writing

$$I_{[\tau_i+\Delta_i,\tau_i+\Delta_{i+1}]}(\tilde{\mathfrak{z}}^a_\delta) \le \sum_{r\in\mathcal{R}} \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \left[\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s)) - \mu^*_r + \mu^*_r \log \frac{\mu^*_r}{\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))}\right] \mathrm{d}s\,.$$

where μ^*_r is given by the multiplicity of reaction r in $\mathcal{E}_j$. We then further divide the sum on $\mathcal{R}$ based on whether the jump rates $\lambda_r(z(t))$ are bounded from below by ε'' from Assumption 2 on the time interval of interest. We denote the jumps whose rates do not satisfy this lower bound by $\mathcal{R}^{(0)}(j_i)$ and write

$$\begin{aligned} I_{[\tau_i+\Delta_i,\tau_i+\Delta_{i+1}]}(\tilde{\mathfrak{z}}^a_\delta) \le & \sum_{r\in\mathcal{R}^{(0)}(j_i)} \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \left[\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s)) - \mu^*_r + \mu^*_r \log \frac{\mu^*_r}{\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))}\right] \mathrm{d}s \\ & + \sum_{r\in\mathcal{R}\setminus\mathcal{R}^{(0)}(j_i)} \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \left[\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s)) - \mu^*_r + \mu^*_r \log \frac{\mu^*_r}{\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))}\right] \mathrm{d}s\,. \end{aligned} \tag{3.19}$$

Then, by compactness given by $I(z) < K$ there exists $C'(K) > 0$ such that the second term is bounded from above by

$$\sum_{r\in\mathcal{R}\setminus\mathcal{R}^{(0)}(j_i)} \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \left[\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s)) - \mu^*_r + \mu^*_r \log \frac{\mu^*_r}{\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))}\right] \mathrm{d}s < |\mathcal{R}| C'(\Delta_{i+1} - \Delta_i)\,.$$

On the other hand, for the first term in (3.19) we have

$$\begin{aligned} & \sum_{r\in\mathcal{R}^{(0)}(j_i)} \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \left[\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s)) - \mu^*_r + \mu^*_r \log \frac{\mu^*_r}{\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))}\right] \mathrm{d}s \\ & \qquad < |\mathcal{R}| C'(\Delta_{i+1} - \Delta_i) + |\mathcal{R}| C'' \int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \log \lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))\,\mathrm{d}s\,, \end{aligned} \tag{3.20}$$

and using Assumption 2 b) and c) we have for some $x \in \mathcal{A}_{j_i}$

$$\lim_{\delta\to 0}\int_{\tau_i+\Delta_i}^{\tau_i+\Delta_{i+1}} \log\lambda_r(\tilde{\mathfrak{z}}^a_\delta(s))\,\mathrm{d}s \le \lim_{\delta\to 0}\int_0^{\beta^i t_\delta} \log\lambda_r(x+sw_{j_i})\,\mathrm{d}s = 0\,.$$

Combining the above upper bounds for each transition we have

$$\lim_{\delta\to 0}\sum_{i=0}^{J} I_{[\tau_i+\Delta_i,\tau_i+\Delta_{i+1}]}(\tilde{\mathfrak{z}}^a_\delta) \le J\lim_{\delta\to 0}\sup_{i\in(0,\dots,J)} I_{[\tau_i+\Delta_i,\tau_i+\Delta_{i+1}]}(\tilde{\mathfrak{z}}^a_\delta) = 0\,. \tag{3.21}$$

Finally, combining (3.17) with (3.18) and (3.21) we obtain the desired result. □

3.4 Proof of LDP in path space

Proof of Proposition 3.1. We conclude the proof by bounding the terms in (3.4). For the first one we have that (3.5) holds by combining (3.8) and Lemma 3.4, for which we have that

$$\lim_{\delta\to 0}\liminf_{v\to\infty}\frac{1}{v}\log\mathbb{P}\left[X^v(t)\in\mathcal{B}_{\delta''/2}(t_\delta w_j)\,,\ \Xi_j(n^v_+,v)\,,\ \sigma_{n^v_+,n_j}>t_\delta\middle|X^v(0)=x^v_0\right]=0\,, \tag{3.22}$$

It remains to show that the second term is bounded by the rate function as in (3.5). We first bound this term as

$$\inf_{y\in\mathcal{B}_{\delta''/2}(x_0+t_\delta w)}\mathbb{P}\left[X^v\in\mathrm{B}_{[t_\delta,T]}(\delta',\mathfrak{z}_\delta)\mid X^v(t_\delta)=y\right]\ge$$
$$\inf_{a\in\mathcal{B}_{\delta''/2}(0)}\mathbb{P}\left[X^v\in\mathrm{B}_{[t_\delta,T]}(\delta'/2,\mathfrak{z}_\delta+a)\mid X^v(t_\delta)=\mathfrak{z}_\delta(t_\delta)+a\right],$$

where we shift the path $\mathfrak{z}_\delta$ of a, but the lower bound is preserved since $\mathrm{B}_{[t_\delta,T]}(\delta'/2,\mathfrak{z}_\delta+a)\subseteq\mathrm{B}_{[t_\delta,T]}(\delta',\mathfrak{z}_\delta)$ for all $a\in\mathcal{B}_{\delta''/2}(0)$. Since paths in the RHS above are uniformly bounded away from $\partial\mathcal{S}$ by Lemma 3.2, rates are uniformly bounded away from 0 on the paths of interest and standard large-deviation bounds (which hold uniformly on $y\in\mathcal{B}_{\delta''/2}(t_\delta\, w_{i_0})$) can be applied. Therefore, defining $\mathcal{N}_\delta := \mathcal{B}_{\delta''/2}(0)\cap\mathrm{span}_{r\in\mathcal{R}}(\gamma^r)$ we bound the second term of (3.4) by

$$\liminf_{v\to\infty}\frac{1}{v}\log\inf_{a\in\mathcal{N}_\delta}\mathbb{P}\Big[X^v\in\mathrm{B}_{[t_\delta,T]}(\delta'/2,\mathfrak{z}_\delta+a)\,|\,X^v(t_\delta)=\mathfrak{z}_\delta(t_\delta)+a\Big] \tag{3.23}$$
$$\ge\inf_{a\in\mathcal{N}_\delta}\Big(-\inf_{z\in\mathrm{B}_{[t_\delta,T]}(\delta'/2,\mathfrak{z}_\delta+a)} I_{[t_\delta,T]}(z)\Big)$$
$$\ge\inf_{a\in\mathcal{N}_\delta}\left(-I_{[t_\delta,T]}(\mathfrak{z}_\delta+a)\right)$$
$$\ge\inf_{a\in\mathcal{N}_\delta}\left(-I_{[0,T]}(\tilde{\mathfrak{z}}^a_\delta)\right). \tag{3.24}$$

Finally, combining Lemma 3.8 with the bound obtained above we obtain that

$$\liminf_{\delta\to 0}\inf_{a\in\mathcal{N}_\delta}\left(-I_{[0,T]}(\tilde{\mathfrak{z}}^a_\delta)\right)\ge -I_{[0,T]}(z) = -I^{x_0}_{[0,T]}(z)\,. \tag{3.25}$$

□

We conclude this section by proving Lemma 3.2

Proof of Lemma 3.2. The finiteness of J results from [SW05, Lemma 3.5]. In particular, one can choose α small enough and $[\tau_{i-1}, \tau_i]$ so that the set $\{\mathcal{B}_\alpha(z(t)) \ , \ t \in [\tau_{i-1}, \tau_i]\}$ is contained in $\mathcal{A}_{j_i}$ for all $i \in (1, \dots, J)$. By exponential tightness we can apply [SW05, Lemma 3.4] to obtain absolute continuity of z on the set $I_{[0,T]}(z) < K$, so that there exists $\tau_- > 0$ with $\inf_{\{z \,:\, I_{[0,T]}(z)<K, i\in\mathbb{N}\}} \tau_i - \tau_{i-1} > \tau_-$. Consequently $J = T/\tau_-$ is finite.

The bound $\sup_{[0,T]} \|z - \mathfrak{z}_\delta\| < 2\delta/3$ follows from the construction (3.2) and our choice of $\xi := \min(1, (\kappa''/3)^{J+1}/3, \varepsilon)$. Indeed for the time interval $[0, t_\delta]$ we have

$$\sup_{t\in[0,t_\delta]} \|x_0 + w_j t - z(t)\| \le t_\delta + \omega_z(t_\delta) \le \frac{2\delta}{3}\,,$$

where ω_z is the (subadditive) modulus of continuity of z. To extend this estimate beyond t_δ we note that

$$\sup_{t\in[0,T]} \|\mathfrak{z}_\delta(t) - z(t)\| < \sum_{k=0}^{J} \beta^k t_\delta + \omega_z(\beta^k t_\delta) \le 2\xi\delta \sum_{k=0}^{J} \beta^k\,. \tag{3.26}$$

Then, by our choice $\beta = 3/\kappa''$ and since

$$\sum_{l=0}^{k} \left(\frac{3}{\kappa''}\right)^l = \frac{1-(3/\kappa'')^{k+1}}{1-3/\kappa''} \le \frac{\kappa''}{2}(3/\kappa'')^{k+1}\,. \tag{3.27}$$

we see that by boundedness of $k \le J$ and by the definition of ξ we have $\sup_{t\in[0,T]} \|\mathfrak{z}_\delta(t) - z(t)\| < 2\delta/3$.

We now prove that $\bigcup_{t\in[t_\delta,T]} \mathcal{B}_{\kappa_- t_\delta}(\mathfrak{z}_\delta(t)) \cap \partial\mathcal{S} = \varnothing$ by induction on k. For $k = 0$ the claim follows directly by Assumption 2a) for $\delta < \varepsilon'/2$. Then, by (3.27) as the $k+1$-th shift is of length $t_\delta(3/\kappa'')^{k+1}$ we must have that $\mathfrak{z}_\delta(t)$ is at least at distance $t_\delta\kappa_-(1-\kappa''/2)(3/\kappa'')^{k+1} > t_\delta\kappa_-/2(3/\kappa'')^{k+1} > \kappa_- t_\delta$ from $\partial\mathcal{S}$ for $t \in [\tau_{k+1} + \Delta^\beta_{k+1}, \tau_{k+2} + \Delta^\beta_{k+1}]$. Since the initial point of the shift satisfies the required condition by assumption, and that this property is conserved on $[\tau_k + \Delta^\beta_k, \tau_k + \Delta^\beta_{k+1}]$ (*i.e.*, during a shift) by Assumption 2a) we obtain the desired result.

Finally, we show that for every $k \in (1, \dots, J)$ and $a \in \mathbb{R}^d$ with $\|a\| < t_\delta\kappa''/2$ there exists $\tilde{w} \in \mathcal{B}_{\kappa''}(0)$ such that $\mathfrak{z}_\delta(\tau_k + \Delta^\beta_{k+1}) + a = z(\tau_k) + \beta^k t_\delta(w_{i_k} + \tilde{w})$. This follows immediately from (3.27), since we have

$$\begin{aligned}\|\mathfrak{z}_\delta(\tau_k + \Delta^\beta_{k+1}) + a - z(\tau_k) - \beta^k t_\delta w_{i_k}\| &= \|\sum_{l=0}^{k-1} \left(\frac{3}{\kappa''}\right)^l t_\delta w_{i_l} + a\| \le \sum_{l=0}^{k-1} \left(\frac{3}{\kappa''}\right)^l t_\delta + \|a\| \\ &\le \frac{t_\delta\kappa''}{2}\left(1 + (3/\kappa'')^k\right) \le \kappa''\|(3/\kappa'')^k t_\delta w_{i_k}\|\,,\end{aligned}$$

concluding the proof of the lemma. □

4 LDP upper bound

Similar results under slightly more restrictive assumptions are well known *e.g.*, [DEW91, SW95] with jump rates bounded away from 0. A sufficiently general result is available in [PR19], but under assumptions on the initial condition that are not satisfied here. We will sketch the application of the ideas from [PR19] to the setting of this paper.

In order to prove the upper bound, we will temporarily enlarge the state space in order to include the integrated flux of each reaction, i. e. we consider the process $(X^v(t), W^v(t)) \in \mathbb{R}^d \times \mathbb{R}^{|\mathcal{R}|}_{\geq 0}$ with initial condition $(x_0^v, 0)$ and generator:

$$Q^v f(x,w) = \sum_{r\in\mathcal{R}} v\Lambda_r^v(x)(f(x+v^{-1}\gamma^r, w+v^{-1}\delta^r) - f(x,w)),$$

for $\delta_r^r = 1$ and $\delta_s^r = 0$ for all $s \neq r$. It is clear that the marginal distribution of the X^v-coordinate is the distribution of our original process. In the following proposition, we prove a large-deviation upper bound for this process. To shorten notation, let us define for any $w \in \mathbb{R}^{|\mathcal{R}|}$ the vector $\Gamma w \colon = \sum_{r\in\mathcal{R}} \gamma^r w_r \in \mathbb{R}^d$.

Let $C_c^1([0,T);\mathbb{R}^{|\mathcal{R}|})$ be the space of continuous and differentiable compactly supported functions from $[0,T)$ to $\mathbb{R}^{|\mathcal{R}|}$. For $x \in D_u(0,T;\mathcal{S})$, $w \in D_u(0,T;\mathbb{R}^{|\mathcal{R}|}_{\geq 0})$ and $\zeta \in C_c^1([0,T);\mathbb{R}^{|\mathcal{R}|})$ we set

$$G(x,w,\zeta) := -\int_0^T \sum_{r\in\mathcal{R}} \Big(\dot{\zeta}_r(t) w_r(t) + \left[e^{\zeta_r(t)} - 1\right]\lambda_r(x(t))\Big)\,\mathrm{d}t$$

and use this to define a partial rate function

$$\widetilde{\mathcal{J}}_{\mathcal{S}}(x,w) := \begin{cases} \sup_{\zeta\in C_c^1([0,T);\mathbb{R}^{\mathcal{R}})} G(x,w,\zeta) & \text{if } x(t) = x_0 + \Gamma w(t), \quad x(t)\in\mathcal{S} \quad \forall t\in[0,T) \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition 4.1. *Let $\zeta \in C_c^1([0,T);\mathbb{R}^{|\mathcal{R}|})$ and suppose Assumption 1 holds, then $(x,w) \mapsto G(x,w,\zeta)$ is continuous from $D_u(0,T;\mathbb{R}^d \times \mathbb{R}^{|\mathcal{R}|}_{\geq 0})$ to $\mathbb{R}$.*

Proof. We have

$$|G(x,w,\zeta) - G(x',w',\zeta)| \leq \left\|\dot{\zeta}\right\|_\infty \|w - w'\|_\infty T + \left(\mathrm{e}^{\|\zeta\|_\infty} + 1\right)\int_0^T \sum_{r\in\mathcal{R}} |\lambda_r(x(t)) - \lambda_r(x'(t))|\,\mathrm{d}t.$$

We can use the continuity of the λ_r from Assumption 1a), along with the boundedness given by the compactness of $\mathcal{S}$, and apply dominated convergence when $x' \to x$ to see that the second term of our estimate vanishes, as well as the first term when $w' \to w$. □

Proposition 4.2. *Let $\mathcal{K}$ be a closed subset of the space of càdlàg paths $D_u(0,T;\mathbb{R}^d \times \mathbb{R}^{|\mathcal{R}|}_{\geq 0})$, then under Assumption 1*

$$\limsup_v \frac{1}{v}\log\mathbb{P}\left((X^v, W^v)\in\mathcal{K}\right) \leq -\inf_{(x,w)\in\mathcal{K}} \widetilde{\mathcal{J}}_{\mathcal{S}}(x,w).$$

Proof. Assumption 1 implies exponential tightness – see discussion below the statement of Theorem 1 – therefore we may assume that $\mathcal{K}$ is compact.

Fix $\varepsilon \in (0,1)$, then for every $(x,w) \in D_u(0,T;\mathcal{S}\times\mathbb{R}^{|\mathcal{R}|}_{\geq 0})$ satisfying $x = x_0 + \Gamma w$ one can find $\zeta[x,w] \in C_c^1\left([0,T);\mathbb{R}^{\mathcal{R}}\right)$ such that

$$G\left(x,w,\zeta[x,w]\right) \geq \min\Big(\widetilde{\mathcal{J}}(x,w), \varepsilon^{-1}\Big) - \varepsilon$$

and define neighbourhoods in path space

$$\mathcal{G}_\varepsilon(x,w) := \left\{(x',w') : G(x',w',\zeta[x,w]) \geq G(x,w,\zeta[x,w]) - \varepsilon\right\}.$$

We may use Proposition 4.1 to see that the $\mathcal{G}_\varepsilon(x,w)$ are open and so can find a finite cover $\mathcal{G}_\varepsilon(x^i,w^i)$ $i=1,\dots,n$ for $\mathcal{K}$.

Now following [PR19, Thm A.3] and using the fact that the jump rates are bounded over $\mathcal{S}$ (because of Assumption 1a) and compactness of $\mathcal{S}$), we define tilted measures $\mathbb{P}^\zeta$ via mean 1 non-negative martingales so that

$$\frac{1}{v}\log\frac{\mathrm{d}\,\mathbb{P}^\zeta\circ(X^v,W^v)^{-1}}{\mathrm{d}\,\mathbb{P}\circ(X^v,W^v)^{-1}}(x,w) = -\int_0^T\sum_{r\in\mathcal{R}} w_r(t)\dot{\zeta}_r(t)+\Lambda_r^v(x(t))\left(\mathrm{e}^{\zeta_r(t)}-1\right)\mathrm{d}t := G^v(x,w,\zeta).$$

Slightly adapting and simplifying the argument of [PR19, Lemma 4.7]

$$\begin{aligned}
&\frac{1}{v}\log\mathbb{P}\left((X^v,W^v)\in\mathcal{G}_\varepsilon(x,w)\right)\\
&\qquad\le\frac{1}{v}\log\mathbb{P}^{\zeta[x,w]}\left((X^v,W^v)\in\mathcal{G}_\varepsilon(x,w)\right)-\inf_{(x',w')\in\mathcal{G}_\varepsilon(x,w)}G^v(x',w',\zeta[x,w])\\
&\le-\inf_{(x',w')\in\mathcal{G}_\varepsilon(x,w)}\left|G^v(x',w',\zeta[x,w])-G(x',w',\zeta[x,w])\right|-\inf_{(x',w')\in\mathcal{G}_\varepsilon(x,w)}G(x',w',\zeta[x,w]).
\end{aligned}\tag{4.1}$$

The first term on the final line vanishes as $v\to\infty$ by the uniform convergence from Assumption 1b). Thus for $i=1,\dots,n$

$$\frac{1}{v}\log\mathbb{P}\left((X^v,W^v)\in\mathcal{G}_\varepsilon(x^i,w^i)\right)\le-\min\left(\widetilde{\mathcal{J}}(x^i,w^i),\varepsilon^{-1}\right)-2\varepsilon$$

and by the Laplace principle

$$\frac{1}{v}\log\mathbb{P}\left((X^v,W^v)\in\mathcal{K}\right)\le-\min_{i=1,\dots,n}\min\left(\widetilde{\mathcal{J}}(x^i,w^i),\varepsilon^{-1}\right)-2\varepsilon\le\inf_{(x',w')\in\mathcal{K}}\min\left(\mathcal{J}(x',w'),\varepsilon^{-1}\right)-2\varepsilon,$$

which completes the proof as ε can be taken arbitrarily small.

If $\nexists\,(x,w)\in\mathcal{K}$ satisfying $x=x_0+\Gamma w$, then $\limsup_v\frac{1}{v}\log\mathbb{P}\left((X^v,W^v)\in\mathcal{K}\right)\le-\infty$, since, by definition $X^v(t)=X^v(0)+\Gamma W^v(t)$ a.s. for all $t\in[0,T]$. □

Proposition 4.3. *If Assumption 1 holds then $\mathcal{J}=\widetilde{\mathcal{J}}_\mathcal{S}$, where*

$$\mathcal{J}(x,w)=\begin{cases}\sum_{r\in\mathcal{R}}\int_0^T\mathcal{H}(\dot{w}(t)|\lambda(x(t)))\,\mathrm{d}t, & (x,w)\in\mathcal{AC}(0,T;\mathcal{S}\times\mathbb{R}^{|\mathcal{R}|}),\dot{x}=\Gamma\dot{w},x(0)=x_0\\ \infty, & \textit{otherwise}.\end{cases}$$

where $\mathcal{H}$ is defined in (2.5)*.*

Proof. The proof follows [PR19, Prop 3.5]; we assume that $\dot{x}=\Gamma\dot{w}$ and $x(t)\in\mathcal{S}$ a.e. $t\in[0,T)$ throughout. Suppose that $(x,w)\notin\mathcal{AC}(0,T;\mathcal{S}\times\mathbb{R}^{|\mathcal{R}|}_{\ge0})$, then one can find a sequence $\zeta^n\in C^1_\mathrm{c}\left([0,T);\mathbb{R}^{|\mathcal{R}|}\right)$ with $\sup_{n,r,t}|\zeta^n_z(t)|\le1$ but $\lim_{n\to\infty}\int_0^T\sum_{r\in\mathcal{R}}\dot{\zeta}^n_z(t)w(t)\mathrm{d}t=-\infty$ and thus $\mathcal{J}(x,w)\ge\lim_{n\to\infty}G(c,w,\zeta^n)=+\infty$, that is, $\mathcal{J}=\tilde{\mathcal{J}}_\mathcal{S}$ if the path is not absolutely continuous.

One then shows that $\mathcal{J}(x,w)=\widetilde{\mathcal{J}}_\mathcal{S}(x,w)$ for any $(x,w)\in\mathcal{AC}(0,T;\mathcal{S}\times\mathbb{R}^{|\mathcal{R}|})$ using approximation arguments. □

Corollary 4.4. *If Assumption 1 holds then the large-deviation upper bound holds with good rate functional:*

$$\inf_{\substack{w\in W^{1,1}(0,T;\mathbb{R}^{|\mathcal{R}|}_{\geq 0})\\ \dot{x}=\Gamma\dot{w}}} \mathcal{J}(x,w) = I^{x_0}_{[0,T]}(x) = \int_0^T \sup_{\vartheta\in\mathbb{R}^d}\Big[\vartheta\cdot\dot{x}(t) - \sum_{r\in\mathcal{R}}\lambda_r(x(t))\big(\exp(\vartheta\cdot\gamma^r)-1\big)\Big]\,\mathrm{d}t.$$

Proof. The large-deviation upper bound with the rate functional on the left-hand side follows from Propositions 4.2, 4.3 and the contraction principle. For non-absolutely continuous paths both the left-hand and right-hand sides will blow up: The left-hand side since the infimum will be taken over an empty set, and the right-hand by a similar argument as in Proposition 4.3.

Now take an arbitrary $x\in\mathcal{AC}(0,T;\mathbb{R}^d)$. Then

$$\begin{aligned}\inf_{\substack{w\in W^{1,1}(0,T;\mathbb{R}^{|\mathcal{R}|}_{\geq 0})\\ \dot{x}=\Gamma\dot{w}}} \mathcal{J}(x,w) &\geq \underbrace{\int_0^T \inf_{j\in\mathbb{R}^{|\mathcal{R}|}_{\geq 0}:\dot{x}(t)=\Gamma j}\mathcal{H}\big(j(t)|\lambda(x(t))\big)\,\mathrm{d}t}_{=I^{x_0}_{[0,T]}(x)}\\ &= \int_0^T \sup_{\vartheta\in\mathbb{R}^d}\Big[\vartheta\cdot\dot{x}(t) - \sum_{r\in\mathcal{R}}\lambda_r(x(t))\big(\exp(\vartheta\cdot\gamma^r)-1\big)\Big]\,\mathrm{d}t,\end{aligned}$$

where the equality follows from convex duality, pointwise in t. To show that the inequality is in fact an equality, we may assume that the left-hand side is finite. Hence from now on we may assume that $x\in W^{1,1}(0,T;\mathcal{S})$. By Jensen's inequality, any path $j:(0,T)\to\mathbb{R}^{|\mathcal{R}|}_{\geq 0}$ for which $\int_0^T\mathcal{H}\big(j(t)|\lambda(x(t))\big)\,\mathrm{d}t<\infty$ is bounded in $L^1(0,T;\mathbb{R}^{|\mathcal{R}|})$, which shows the first equality.

□

5 Optimality of decay rate in (2.3)

We recall from Remark 2.5 that integrability of the rates necessary to establish the lower bound estimates in Section 3 – but not the upper bound ones in Section 4 – is directly implied by a sufficiently slow decay of the rates (2.3). In this section (and more specifically in Proposition 5.4) we make precise our claim that the range of exponents α given in Remark 2.5 is maximal. In particular, we show that whenever the rates of jumps necessary to escape the degenerate set decay too fast (satisfying a consition similar to (2.3) for $\alpha\geq 1$), the rate function for the upper bound diverges for any $y\in\mathcal{AC}(0,T;\mathcal{S})$ with $y(0)\in\partial\mathcal{S}$ and $y(t)\in\mathcal{S}\setminus\partial\mathcal{S}$ for a $t>0$. We know from Corollary 4.4 that, under Assumption 1, for any $y\in\mathcal{AC}(0,T;\mathcal{S})$ and any $\delta>0$:

$$\limsup_{v\to\infty}\frac{1}{v}\log\mathbb{P}_{x_0^v}(X^v\in\overline{\mathrm{B}_{[0,T]}(\delta,y)}) \leq -\inf_{x\in\overline{\mathrm{B}_{[0,T]}(\delta,y)}}\int_0^T \ell(x(s),\dot{x}(s))\mathrm{d}s, \tag{5.1}$$

where we recall that ℓ is defined as

$$\ell(x,y) = \sup_{\vartheta\in\mathbb{R}^d}\vartheta\cdot y - \sum_{r\in\mathcal{R}}\lambda_r(x)\left(\exp\left(\vartheta\cdot\gamma_r\right)-1\right).$$

We start the discussion of this problem with some examples as to capture the idea of our strategy in a simple setting.

Example 5.1. *Recall Example 2.4, where (2.3) does not hold and the upper bound is easily seen to diverge. Given the generator in* (2.1)*, the integrand on the right-hand side of the* LDP *upper bound in* (5.1) *reads*

$$\ell(x,y) = \sup_{\vartheta\in\mathbb{R}^d} \left\{ \vartheta y - e^{-1/x}(e^{\vartheta}-1) \right\} \geq y \log y e^{1/x} - (y - e^{-1/x}),$$

where in the inequality we have chosen $\vartheta(x,y) = \log y e^{1/x}$. Then, the LDP *rate function can be bounded as follows*

$$I^0_{[0,T]}(z) \geq \int_0^T z'(t)\log(z'(t)e^{1/z(t)}) - (z'(t) - e^{-1/z(t)})\mathrm{d}t \geq \int_0^T z'(t)\log z'(t) + z'(t)/z(t) - z(t)\,\mathrm{d}t,$$

for any $z \in \mathcal{AC}(0,T;\mathbb{R}_{\geq 0})$ with $z(0)=0$. Using that $x\log x > -1$ is continuous at 0 we have, for every bounded path z with bounded derivative, that

$$I^0_{[0,T]}(z) \geq -1 - z(1) + \int_0^1 z'(t)/z(t)\,\mathrm{d}t \geq -1 - z(T) + \int_0^{z(T)} x^{-1}\,\mathrm{d}x = +\infty.$$

In order to proceed and generalize the example above, we discuss two further examples highlighting the appropriate way to negate the assumptions in Remark 2.5.

Example 5.2. *Consider a system defined on $\mathcal{S}_v = (v^{-1}\mathbb{N}_0)^2$ with two jumps, $\gamma_1 = (0,1)$, $\gamma_2 = (1,0)$ and corresponding rates $\Lambda^v_1(x) = \lambda_1(x) = \mathbb{1}\{x_1 \geq 1\}(x_1 - 1)$, $\Lambda^v_2(x) = \lambda_2(x) = 1$. For the initial condition $x_0 = 0$ the law of large numbers [Kur72] shows that the paths of X^v concentrate around $(y^*_1, y^*_2)(t) = (t, \mathbb{1}\{t>1\}(t-1)^2/2)$. In particular, this implies the existence of paths $y(t) \in \mathcal{AC}(0,T;\mathcal{S})$ with $y(0) \in \partial\mathcal{S}$ and $y(t) \in \mathcal{S}\setminus\partial\mathcal{S}$ but having a finite large deviations cost for a system violating (2.3).*

To discuss the optimality of the interval for the parameter α governing the *local* decay of rates in (2.3) we avoid considering macroscopic behaviors like the one highlighted in the example above and we restrict our attention to paths that do not leave the set $\mathcal{A}_j$ in the time interval of interest, keeping j fixed throughout this section. For the same reason, to negate (2.3) we consider jumps whose rates decay faster than $\exp[-k \cdot \mathrm{dist}(x,\partial\mathcal{A}_j)^{-1}]$ *uniformly* in x in $\mathcal{A}_j$ for a $k > 0$. These jumps belong to the set

$$\mathrm{FAST}_{\mathcal{R},j} := \left\{ r \in \mathcal{R} \;:\; \lim_{\varrho\to 0} \varrho \left(\sup_{z\in\mathcal{A}_j:\, \mathrm{dist}(z,\partial\mathcal{A}_j)<\varrho} \log\lambda_r(z) \right) < 0 \right\}, \tag{5.2}$$

with $\mathrm{dist}(z,\partial\mathcal{A}_j) = \inf_{x\in\partial\mathcal{A}_j}\|z-x\|$. While this set is not, in general, the complement of the jumps whose rates satisfy (2.3), it allows to capture, at least locally, those whose decay is more rapid (in terms of α) than (2.3).

We further notice that the existence of a single reaction $r \in \mathrm{FAST}_{\mathcal{R}}$ may still result in a finite cost for paths escaping $\partial\mathcal{A}_j$, as the following example shows.

Example 5.3. *Consider a system defined on $\mathcal{S}_v = (v^{-1}\mathbb{N}_0)^2$ with two jumps, $\gamma_1 = (-1,1)$ and $\gamma_2 = (1,1)$ and corresponding rates $\lambda_1(x) = x_1 e^{-1/x_2}$, $\lambda_2(x) = 1$. It is clear that this system satisfies a* LDP *with any sequence of initial conditions, see [SW05]. However, approaching the set $\{x \in \mathbb{R}^2 \,:\, x_2 = 0\}$, upon choosing $w_j = (1,0)$ we see that this system does not satisfy (2.3). We are, however, able to* choose $w_j = (1,1)$ *so that* with such choice of w_j *(2.3) holds for all $r \in \mathcal{E}_j$.*

In light of the above example, the statement we seek to negate is the existence of vectors w_j and corresponding $\mathcal{E}_j$ stated in Assumption 2 such that $\mathcal{E}_j \cap \mathrm{FAST}_{\mathcal{R},j} = \varnothing$. To do so, we fix a (non-empty) $\mathcal{S}$ and a limiting point $x_0 \in \partial\mathcal{A}_j$, for $j \in \mathcal{I}^{\mathrm{bd}}$, assuming throughout that $\partial\mathcal{A}_j$ is a $(d-1)$-dimensional hyperplane to simplify the notation of the proof. In this way, we define $T_x\mathcal{A}_j = \{y \in \mathrm{span}_{r\in\mathcal{R}}(\gamma^r) \colon y \cdot n_x \geq 0\}$, with n_x inward normal to the boundary $\partial\mathcal{A}_j$ in x. Assuming that there is no vector $w_j \in \mathrm{span}_{r\in\mathcal{R}}(\gamma^r)$ that is a sum of jumps with rates decaying slow enough means that $\forall x \in \partial\mathcal{A}_j$

$$\mathrm{Co}_x\left(\{\gamma^r \ : r \in (\mathcal{R}\setminus \mathrm{FAST}_{\mathcal{R},j})\}\right) \cap T_x\mathcal{A}_j = \varnothing\,, \tag{5.3}$$

where $\mathrm{Co}_x(A)$ is the the convex cone defined by the set of vectors A with origin x.

Note that in this way we are building a class of processes where the jumps r pointing in the interior of the domain (and therefore useful to escape the boundary) necessarily belong to $\mathrm{FAST}_{\mathcal{R},j}$.

Proposition 5.4. *Assume that (5.3) holds, then for every $y \in \mathcal{AC}(0,T;\mathcal{S})$ with $y(0) = x_0 \in \partial\mathcal{A}_j$ and such that there exists $t_1 \in (0,T)$ with $y(t_1) \in \mathcal{A}_j \setminus \partial\mathcal{A}_j$ and $\inf\{t \in (0,T) \ : \ z(t) \notin \mathcal{A}_j\} > t_1$ it holds that $I^{x_0}_{[0,T]}(z) = \infty$.*

Proof. Recalling the structure in (5.1) of the LDP upper bound we notice that, in order to show that the rate function is infinite for any path $y \in \mathcal{AC}(0,T;\mathcal{S})$ as above, it is sufficient to find a $\vartheta(t,y)$ such that

$$\int_0^T \left[\vartheta(s,y(s))\cdot\dot{y}(s) - \sum_{r\in\mathcal{R}} \lambda_r(y(s))\left(\exp\left[\vartheta(s,y(s))\cdot\gamma_r\right] - 1\right)\right]\mathrm{d}s = +\infty\,. \tag{5.4}$$

By assumption, there exists $t_0 < t_1 \in [0,T]$ such that $y(t_0) \in \partial\mathcal{A}_j$ and $y(t) \notin \partial\mathcal{A}_j$ for all $t \in (t_0,t_1)$. We now aim to express $\int_{t_0}^{t_1}\vartheta(t,y(t))\cdot\dot{y}(t)\,\mathrm{d}t$ as an exact integral of some potential for which $\Phi(x_0) = -\infty$, while we choose $\vartheta(t,y(t)) = 0$ for $t \in [0,t_0] \cup [t_1,T]$, such that the integral in (5.4) vanishes on that interval. More specifically, following for example [HPST20], for a $\kappa > 0$ we take $\vartheta = \kappa\nabla\Phi(y)$ so that

$$I(y) \geq \kappa\Phi(y(t_1)) - \kappa\Phi(y(t_0)) - \sum_{r\in\mathcal{R}}\int_{t_0}^{t_1}\lambda_r(y(t))e^{\kappa\gamma_r\cdot\nabla\Phi(y)}\,\mathrm{d}t + \underbrace{\sum_{r\in\mathcal{R}}\int_{t_0}^{t_1}\lambda_r(y(t))\,\mathrm{d}t}_{\geq 0}\,.$$

The missing step is therefore to choose Φ and tune κ such that $\Phi(y(t_0)) = -\infty$ and $\sum_{r\in\mathcal{R}}\int_{t_0}^{t_1}\lambda_r(y(t))e^{\kappa\gamma_r\cdot\nabla\Phi(y)}\,\mathrm{d}t$ is bounded. For the choice $\Phi(y) := \log(n_{x_0}\cdot(y-x_0))$ we have:

$$\sum_{r\in\mathcal{R}}\int_{t_0}^{t_1}\lambda_r(y_t)\exp\left[\kappa\gamma_r\cdot\nabla\Phi(y)\right]\,\mathrm{d}t = \sum_{r\in\mathcal{R}}\int_{t_0}^{t_1} e^{\log\lambda_r(y(t)) + \kappa(n_x\cdot\gamma_r)/(n_{x_0}\cdot(y(t)-x_0))}\,\mathrm{d}t\,.$$

With this choice, since $y \in \mathcal{AC}(0,T;\mathcal{S})$, we have that $n_{x_0}\cdot(y(t)-x_0) \geq 0$ for $t \in (t_0,t_1)$. We can split the jumps in the set $\mathcal{R}_+ := \{r \in \mathcal{R} : \gamma^r \in T_{x_0}\mathcal{A}_j\}$ and $\mathcal{R}_- = \mathcal{R}\setminus\mathcal{R}_+$. For the latter class of jumps we have $n_{x_0}\cdot\gamma_r \leq 0$ and

$$\sum_{r\in\mathcal{R}_-}\int_{t_0}^{t_1}\lambda_r(y(t))e^{\kappa\frac{n_{x_0}\cdot\gamma_r}{n_{x_0}\cdot(y(t)-x_0)}}\,\mathrm{d}t < +\infty\,,$$

since the argument of each integral is bounded on (t_0,t_1). On the other hand, we handle the terms coming from the former class of jumps $\mathcal{R}_+$ – the ones pushing the process in the interior of $\mathcal{S}$ – using that $\mathcal{R}_+ \subseteq \mathrm{FAST}_{\mathcal{R},j}$. Therefore we can tune κ such that

$$\lim_{t\to 0}(n_{x_0}\cdot(y(t)-x_0))\log\lambda_r(y_t) + \kappa(n_{x_0}\cdot\gamma_r) \leq 0 \qquad \forall r \in \mathcal{R}_+\,,$$

ensuring that $\sum_{r\in\mathcal{R}_+}\int_{t_0}^{t_1}\lambda_r(y(t))\exp\left[\kappa\frac{n_{x_0}\cdot\gamma_r}{n_{x_0}\cdot(y(t)-x_0)}\right]\,\mathrm{d}t<+\infty$. This proves (5.4). □

6 List of symbols

$\mathcal{R}$	finite set of jumps/reactions	Subsec. 1.1
Λ_r^v, λ_r	microscopic and macroscopic jump rates	Subsec. 1.1 & Ass. 1, 2
$\gamma^r \in \mathbb{R}^d$	jump vectors	Subsec. 1.1 & Ass. 2
$\mathcal{S}_v, \mathcal{S}, \partial\mathcal{S} \subset \mathbb{R}^d$	reachable points and boundary/degenerate set	Sec. 2
$\mathcal{A}_i, \partial\mathcal{A}_i \subset \mathbb{R}^d$	covering of $\mathcal{S}$	Sec. 2
$\mathcal{I}, \mathcal{I}^{\mathrm{bd}}$,	index sets for covering	Sec. 2
$\mathcal{B}_\varrho(x) \subset \mathbb{R}^d$	euclidean ball of radius ϱ and center x	Subsec. 2.1
$D_u(0,T;\mathbb{R}^d)$ $\big(D_s(0,T;\mathbb{R}^d)\big)$	càdlàg functions with uniform (Skorohod) topology	Subsec. 2.2
$\mathcal{AC}(0,T;\mathbb{R}^d)$ $\big(\mathcal{AC}(0,T;\mathcal{S})\big)$	absolutely continuous functions (restricted to $\mathcal{S}$)	Subsec. 2.2
$\mathrm{B}_{[0,T]}(\varrho, z)$	ball of radius ϱ and center z in $D_u(0,T;\mathbb{R}^d)$	Subsec. 2.2
$I_{[0,T]}, I_{[0,T]}^{x_0}$	large-deviation action and rate functional	Eqns. (2.4), (2.6)
$\mathcal{H}(\mu\|\lambda), \ell(x,y)$	relative entropy and Lagrangian	Eqn. (2.5), (3.16)
$\mathcal{J}, \widetilde{\mathcal{J}}_{\mathcal{S}}$	flux large-deviation functional and dual form	Section 4
$\mathcal{E}_j \subset \mathcal{R}, w_j \in \mathbb{R}^d, \alpha_j > 0$	escape sequence, vector and normalisation	Ass. 2 a), b)
$\varepsilon, \varepsilon', \kappa_j, \kappa_- > 0,$	escape parameters	Ass. 2 a)
$\varepsilon'', \kappa'' > 0$	monotonicity range	Ass. 2 d)
$z, \mathfrak{z}_\delta$	target and approximated path	Subsec. 3.1
$t_\delta, \xi, \beta > 0, \omega_z$	path shift parameters, modulus of continuity	Eqns. (3.2),(3.3)
$\delta', \delta'' > 0$	neighborhood parameters of shifted path	Subsec. 3.1

References

[ACKN20] David F. Anderson, Daniele Cappelletti, Jinsu Kim, and Tung D. Nguyen. Tier structure of strongly endotactic reaction networks. *Stochastic Processes and their Applications*, 130(12):7218 – 7259, 2020.

[ADE18a] Andrea Agazzi, Amir Dembo, and Jean-Pierre Eckmann. Large deviations theory for markov jump models of chemical reaction networks. *The Annals of Applied Probability*, 28(3):1821–1855, 2018.

[ADE18b] Andrea Agazzi, Amir Dembo, and Jean-Pierre Eckmann. On the geometry of chemical reaction networks: Lyapunov function and large deviations. *Journal of Statistical Physics*, 172(2):321–352, 2018.

[Big04] JD Biggins. Large deviations for mixtures. *Electronic Communications in Probability*, 9:60–71, 2004.

[DEW91] Paul Dupuis, Richard S. Ellis, and Alan Weiss. Large deviations for markov processes with discontinuous statistics, i: General upper bounds. *Ann. Probab.*, 19(3):1280–1297, 07 1991.

[DG87] D.A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[DM04] Pierre Del Moral. Feynman-kac formulae. In *Feynman-Kac Formulae*, pages 47–93. Springer, 2004.

[DRW16] Paul Dupuis, Kavita Ramanan, and Wei Wu. Large deviation principle for finite-state mean field interacting particle systems. Technical Report 1601.06219, arXiv, 2016.

[DZ87] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic modelling and applied probability*. Springer, New York, NY, USA, 2nd edition, 1987.

[Fen94] S. Feng. Large deviations for empirical process of mean-field interacting particle system with unbounded jumps. *The Annals of Probability*, 22(4):1679–2274, 1994.

[FK06] J. Feng and T.G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical surveys and monographs*. American Mathematical Society, Providence, RI, USA, 2006.

[FW12] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer, 2012.

[GVE19] Tobias Grafke and Eric Vanden-Eijnden. Numerical computation of rare events via large deviation theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6):063118, 2019.

[HPST20] Bastian Hilder, Mark A Peletier, Upanshu Sharma, and Oliver Tse. An inequality connecting entropy distance, fisher information and large deviations. *Stochastic Processes and their Applications*, 130(5), 2020.

[Kor97] W. Kordecki. Reliability bounds for multistage structures with independent components. *Statistics & Probability Letters*, 34(1):43 – 51, 1997.

[Kur70] T.G. Kurtz. Solutions of ordinary differential equations as limits of pure jump processes. *Journal of Applied Probability*, 7(1):49–58, 1970.

[Kur72] T. G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[Lé95] C. Léonard. Large deviations for long range interacting particle systems with jumps. *Annales de l'Institut Henri Poincaré, section B*, 31(2):289–323, 1995.

[Mie16] A. Mielke. *On Evolutionary Γ-Convergence for Gradient Systems*, pages 187–249. Springer International Publishing, Cham, Swiss, 2016.

[MPPR15] A. Mielke, R.I.A. Patterson, M.A. Peletier, and D.R.M. Renger. Non-equilibrium thermodynamic principles for nonlinear chemical reactions and systems with coagulation and fragmentation. *WIAS Preprint*, 2165, 2015.

[PR19] Robert I. A. Patterson and D. R. Michiel Renger. Large Deviations of Jump Process Fluxes. *Mathematical Physics, Analysis and Geometry*, 22(3):21, Sep 2019.

[PSK16] Etienne Pardoux and Brice Samegni-Kepgnou. Large deviation principle for poisson driven sdes in epidemic models. *arXiv preprint arXiv:1606.01619*, 2016.

[PSK17] Etienne Pardoux and Brice Samegni-Kepgnou. Large deviation principle for epidemic models. *Journal of Applied Probability*, 54(3):905–920, 2017.

[SW95] Adam Shwartz and Alan Weiss. *Large deviations for performance analysis*. Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[SW05] Adam Shwartz and Alan Weiss. Large deviations with diminishing rates. *Math. Oper. Res.*, 30(2):281–310, 2005.

[WRVE04] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Communications on pure and applied mathematics*, 57(5):637–656, 2004.