Electron. Commun. Probab. **22** (2017), no. 42, 1–11.
DOI: 10.1214/17-ECP70
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Signature inversion for monotone paths

Jiawei Chang[*] Nick Duffield[†] Hao Ni[‡] Weijun Xu[§]

Abstract

The aim of this article is to provide a simple sampling procedure to reconstruct any monotone path from its signature. For every N, we sample a lattice path of N steps with weights given by the coefficient of the corresponding word in the signature. We show that these weights on lattice paths satisfy the large deviations principle. In particular, this implies that the probability of picking up a "wrong" path is exponentially small in N. The argument relies on a probabilistic interpretation of the signature for monotone paths.

Keywords: signature; inversion; monotone paths.
AMS MSC 2010: 60.
Submitted to ECP on February 8, 2017, final version accepted on June 23, 2017.

1 Introduction

1.1 The signature of a path

A path γ is a continuous map from a fixed interval J into a normed vector space $(V, \|\cdot\|_V)$. The length of γ is defined by

$$\|\gamma\| := \sup_{\mathcal{D}(J)} \sum_j \|\gamma(u_{j+1}) - \gamma(u_j)\|_V,$$

where the supremum is taken over all dissections $\mathcal{D}(J) = \{u_j\}$ of the interval J. We say γ has bounded variations if $\|\gamma\| < +\infty$. In what follows, we consider $V = \mathbf{R}^d$. Note that although the choice of the norm on $\mathbf{R}^d$ affects the actual length of γ, it does not affect whether γ has bounded variations or not.

Let $\{e_1, \dots, e_d\}$ denote the standard basis of $\mathbf{R}^d$. For every integer $n \geq 0$, a word of length n is an ordered sequence of n letters from the set $\{e_1, \dots, e_d\}$ (with repetition allowed). We use $|w|$ to denote the length of the word w, that is, the number of letters consisting of the word. For two words $w_1 = e_{i_1} \dots e_{i_n}$ and $w_2 = e_{j_1} \dots e_{j_m}$, their concatenation $w_1 * w_2$ is a new word of length $n+m$ defined by

$$w_1 * w_2 = e_{i_1} \dots e_{i_n} e_{j_1} \dots e_{j_m}.$$

We use $\emptyset$ to denote the empty word, which is the unique word of length 0. The signature of a bounded variations path is defined as follows.

[*]University of Oxford, UK. E-mail: jiawei.chang@maths.ox.ac.uk
[†]Texas A&M University, US. E-mail: duffieldng@tamu.edu
[‡]University College London, UK. E-mail: h.ni@ucl.ac.uk
[§]University of Warwick, UK. E-mail: weijun.xu@warwick.ac.uk

Definition 1.1. Let $\gamma : [0,1] \to \mathbf{R}^d$ be a continuous path of bounded variations. For every integer $n \geq 0$ and every word $w = e_{i_1} \ldots e_{i_n}$, let

$$C_\gamma(w) = \int_{0<u_1<\cdots<u_n<1} d\gamma^{i_1}(u_1) \cdots d\gamma^{i_n}(u_n),$$

where γ^i is the i-th component of γ. The signature of γ is the formal power series

$$X(\gamma) = \sum_{n=0}^{+\infty} \sum_{|w|=n} C_\gamma(w) \cdot w,$$

where the second sum is taken over all words of length n, and we have set $C_\gamma(\emptyset) = 1$ by convention.

We call the collection $\{C_\gamma(w) : |w| = n\}$ the n-th level coefficients in the signature. The signature is a definite integral over the fixed interval where γ is defined. Changing the parametrisation or the size of the interval does not change the signature of γ. For bounded variation path γ, one can re-parametrise at constant speed (or natural parametrisation) so that $\|\dot{\gamma}\| \equiv L$ where the length is measured with respect to the given norm on $\mathbf{R}^d$. In particular, under natural parametrisation, γ is Lipschitz, and hence we can write

$$C_\gamma(w) = \int_{0<u_1<\cdots<u_n<1} \dot{\gamma}^{i_1}(u_1) \cdots \dot{\gamma}^{i_n}(u_n) du_1 \cdots du_n.$$

The signature contains important information about the path. For example, the collection $\{C_\gamma(w) : |w| = 1\}$ reproduces the increment of the path, and the second level coefficients $\{C_\gamma(w) : |w| = 2\}$ represents the (signed) areas enclosed by the projection of the path on $e_i - e_j$ planes.

It was proved by Hambly and Lyons ([5]) that bounded variation paths are uniquely determined by their signatures up to tree-like pieces.[1] In [6], Lyons and one of the authors developed a procedure based on the use of symmetrisation that enables one to reconstruct every $\mathcal{C}^1$ path (when at natural parametrisation) from its signature. The purpose of this article is to give a significant simplification of the reconstruction procedure in the case when γ is monotone.

1.2 Monotone paths and the main result

From now on, we fix our path $\gamma : [0,1] \to \mathbf{R}^d$ that is monotone. If γ is decreasing in any of its components, can we can reflect γ^i to $-\gamma^i$ in that component and the corresponding change in the signature is immediate according to Definition 1.1. Thus, we can assume without loss of generality that γ is monotonically increasing so that

$$\dot{\gamma}^i(t) \geq 0, \qquad \forall t \in [0,1], \quad \forall i = 1, \ldots, d.$$

We also equip $\mathbf{R}^d$ with the ℓ^1 norm, so the length of a monotone path is then simply the sum of all its first level coefficients in the signature. Thus, we can assume without loss of generality that $L = 1$; otherwise one can just simply recover L using the first term in the signature and rescale the path by L^{-1} so that the new path has length 1. Finally, since the signature is invariant under re-parametrisation, we also assume γ is parametrised at unit speed so that

$$\sum_{i=1}^{d} \dot{\gamma}^i(u) = 1, \qquad \forall u \in [0,1]. \tag{1.1}$$

[1] Roughly speaking, two paths α and β are tree-equivalent if the path $\alpha * \beta^{-1}$, obtained by running α first and then β backwards, is a "null-path" in the sense that all the trajectories cancel out themselves. Please refer to [4] for a precise definition.

Also, since we fix our path γ throughout the article, we will omit the letter γ and simply write $C(w)$ for the signature of γ.

Since γ is monotonically increasing, we have $C(w) \geq 0$ for every word w. A direct computation then gives

$$\sum_{|w|=N} C(w) = \frac{L^N}{N!} = \frac{1}{N!}$$

for every integer N, where we have also used the assumption that $L = 1$. This suggests that for every N, the quantities $\{N!C_\gamma(w) : |w| = N\}$ constitute a probability measure on the words of length N, giving each word w with $|w| = N$ the "probability" $N!C(w)$. Now, for every word with length N, we associate to it a lattice path X_N with step size $\frac{1}{N}$ such that X_N is a monotone lattice path parametrised at unit speed, and moves in exactly the same direction as the word w. More precisely, if $w = e_{i_1} \cdots e_{i_N}$, then we define the path X_N associated to w by

$$X_N = \frac{1}{N}\big(e_{i_1} * \cdots * e_{i_N}\big),$$

and we equip it with natural parametrisation. Here, "*" denotes the concatenation of two paths, and we have had an abuse of the notation e_{i_k} also to denote the one-step lattice path moving in the e_{i_k} direction. Now, for every $N \geq 0$, we assign the N-step paths $\{X_N : |w| = N\}$ "probabilities" $N!C(w)$. This gives us a sequence of laws on the space of lattice paths. The main result of our article is the following.

Theorem 1.2. The laws on $\{X_N\}$ above satisfies a large deviations principle on the space of continuous function from $[0,1]$ to $\mathbf{R}^d$.

The above theorem implies that one can reconstruct any monotone path from its signature by sampling directly from the lattice paths with weights given by the corresponding terms in the signature. More precisely, for fixed large N, one "samples" a lattice path $\mathcal{S}_N w$ according to the "probabilities" $\{C(w) : |w| = N\}$. The large deviations principle for these laws in Theorem 1.2 then ensures that the chance of picking a wrong lattice path is exponentially small in N.

The proof of this theorem relies on a probabilistic interpretation of the signature of monotone paths. Once this observation is made, the rest follows directly from standard large deviations techniques. On the other hand, unfortunately, the rate function for the LDP for $\{X_N\}$ does not have a closed form. However, an observation by [3] suggests that we can add another random process T_N (to be defined below) to $\{X_N\}$, so that the pair (X_N, T_N) satisfies LDP with a rate function of closed form. These will be made more precise in Section 2.2 below.

2 Sampling path large deviations

2.1 Probabilistic interpretation

We first give the probabilistic interpretation of the signature of a monotone path. Let $\gamma : [0,1] \to \mathbf{R}^d$ be a monotone path parametrised at unit speed with respect to ℓ^1 norm in the sense of (1.1), and it has length 1 under this assumption. We can associate it with a probability measure on random lattice paths in the following way. Consider d independent Poisson processes run simultaneously on the time interval $[0,1]$ respectively. The intensity for the i-th coordinate component of this Poisson process is $\dot{\gamma}^i(t)$. Let $W(t)$ be the word of ordered letters that arrive up to time t. For example, if at times $0 < u_1 < \cdots < u_5 < t$, the letters e_3, e_2, e_2, e_1, e_3 arrives, then

$$W(t) = e_3 e_2 e_2 e_1 e_3, \qquad W(v) = e_3 e_2, \quad v \in [u_2, u_3).$$

One can make the process $W(\cdot)$ into a lattice path in the following way. Suppose the arrival times are τ_j for $j = 1, 2, \ldots$ with the convention that $\tau_0 = 0$, then W can be defined as a lattice path by setting $W(\tau_0) = 0$, and

$$W(t) = W(\tau_j) + (t - \tau_j)e_{i_j},$$

where e_{i_j} is the arriving letter at time τ_j, and the right hand side above should be understood as the sum of two d-dimensional vectors. We have thus associated to γ a random lattice path W.

We are interested in the laws of W conditional on the total number of arrivals up to time 1. Thus, we let $\mathcal{N}(t)$ be the process counting the total number of arrivals up to time t. Since γ is parametrised at unit speed, $\mathcal{N}(t)$ is a homogeneous Poisson process on $[0, 1]$ with intensity 1.

Now, we condition on the event $\mathcal{N}(1) = N$, that is, there are totally N arrivals up to time 1 (when the path runs out). Let

$$\mathbf{P}^N(\cdot) := \mathbf{P}\big(\cdot\,|\mathcal{N}(1) = N\big)$$

denote the conditional probability. Thus, for every word w with $|w| = N$, with the abuse of notation that W denoting the word generated by the processes, we have precisely the relation

$$C(w) = \frac{1}{N!}\mathbf{P}^N\big(W = w\big) = \frac{1}{N!}\mathbf{P}\big(W = w|\mathcal{N}(1) = N\big). \tag{2.1}$$

This is the probabilistic interpretation of the signature for monotone paths.

2.2 Lattice path sampling and large deviations

From now on, we always condition on $\mathcal{N}(1) = N$. Let $W(t)$ be the random lattice path generated by the conditional Poisson process, and $W_N(t) = \frac{1}{N}W(t)$. Thus, under $\mathbf{P}^N$, every realisation of W_N is a lattice path with step size $\frac{1}{N}$ and total length 1 (N steps).

It should be clear that $W_N \to \gamma$ in probability in the space of continuous paths with uniform topology, although we do not know any reference explicitly stating that. Our aim is to show the large deviations principles for the processes W_N and its random time change version. For this, we first introduce the proper function spaces that the processes live in. Let $\mathcal{C}$ denote the set of continuous functions from $[0, 1]$ to $\mathbf{R}$ with the uniform topology, and let $\mathcal{C}^d$ be d copies of $\mathcal{C}$. Also, we let

$$\mathcal{A} = \Big\{\psi : [0, 1] \to \mathbf{R} \text{ s.t. } \psi(0) = 0, \ \ \psi \text{ continuous and non-decreasing}\Big\},$$

and

$$\mathcal{A}_a^d = \Big\{\boldsymbol{\psi} = (\psi^1, \ldots, \psi^d) : \psi^i \in \mathcal{A}, \sum_{i=1}^d \psi^i(1) = a\Big\}.$$

It is clear that $\psi \in \mathcal{A}$ implies ψ is absolutely continuous with $\dot{\psi} \geq 0$. For the set $\mathcal{A}_a^d$, we will mainly use it for $a = 1$. Note that here and below, we will use plain letters for scalars (or scalar-valued functions) and boldface letters for vectors. For example, we use $\boldsymbol{\psi}$ denote the $\mathbf{R}^d$ valued function in $\mathcal{A}_a^d$, while ψ^i denote its one-dimensional components.

We now define the function $I : \mathbf{R}^+ \times \mathbf{R}^+ \to \mathbf{R}$ to be

$$I(x, y) = x \log(x/y), \qquad x, y > 0. \tag{2.2}$$

Our first aim is to show that the conditional laws

$$\mu_N = \mathcal{L}\big(W(\cdot)|\mathcal{N}(1) = N\big)$$

obey a large deviations principle for processes. In order to derive the rate function, we need the following lemma for its finite dimensional approximations.

Lemma 2.1. Let $k \geq 1$. For every $\{0 = u_0 < u_1 < \cdots < u_k = 1\}$, the conditional laws

$$\mathcal{L}\big(W_N(u_1) - W_N(u_0), \ldots, W_N(u_k) - W_N(u_{k-1}) | \mathcal{N}(1) = N\big)$$

satisfy the large deviations principle with scale N and good rate function

$$\mathcal{I}_k(\mathbf{z}) = \sum_{i=1}^{d} \sum_{j=1}^{k} (u_j - u_{j-1}) \cdot I\left(\frac{z_j^i - z_{j-1}^i}{u_j - u_{j-1}}, \frac{\gamma_{u_j}^i - \gamma_{u_{j-1}}^i}{u_j - u_{j-1}}\right)$$

when $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_k) \in (\mathbf{R}^d)^k$ satisfies $\mathbf{z}_j \geq 0$, and

$$\sum_{i=1}^{d} z_k^i = 1, \qquad z_j^i \leq z_{j+1}^i, \qquad \forall i = 1, \ldots, d, \;\; j = 1, \ldots, k,$$

where we have used the convention that $\mathbf{z}_0 = \mathbf{0}$. Otherwise $\mathcal{I}_k(\mathbf{z}) = \infty$.

Proof. Fix $k \geq 1$ and $0 = u_0 < \cdots < u_k = 1$, and let Λ_N be the logarithmic moment-generating function of the multi-vector $\big(W_N(u_1) - W_N(u_0), \ldots, W_N(u_k) - W_N(u_{k-1})\big)$, conditioned on $\mathcal{N}(1) = N$. Here, each component is a d-dimensional vector, and this should be understood as a random vector in $(\mathbf{Z}/N)^{dk}$. Then, the conditional distribution (on $\mathcal{N}(1) = N$) of this random vector is precisely multinomial with N trials and probabilities

$$p_j^i = \gamma_{u_j}^i - \gamma_{u_{j-1}}^i.$$

Here, p_j^i denotes the probability of the outcome of the trial being "$W_N^i(u_j) - W_N^i(u_{j-1})$". Also, the p_j^i's are already normalised as a probability since we have assumed $L = 1$. Hence, for $\boldsymbol{\theta} = \{\theta_j^i\} \in \mathbf{R}^{dk}$, we have

$$\begin{aligned}\frac{1}{N}\Lambda_N(N\boldsymbol{\theta}) &= \frac{1}{N}\log \mathbf{E} \exp\left(N \sum_{i,j} \theta_j^i \big(W_N^i(u_j) - W_N^i(u_{j-1})\big)\right) \\ &= \log\left(\sum_{i,j} \big(\gamma_{u_j}^i - \gamma_{u_{j-1}}^i\big) e^{\theta_j^i}\right),\end{aligned}$$

where the second equality follows from the moment-generating function for multinomial distribution, and the sum is taken over the range

$$i = 1, \ldots, d, \qquad j \in 1, \ldots, k.$$

Hence, the sequence $\{\Lambda_n\}$ satisfies the assumption of Gärtner-Ellis theorem ([2, Theorem 2.3.6]). Let

$$\Lambda(\boldsymbol{\theta}) := \log\left(\sum_{i,j} \big(\gamma_{u_j}^i - \gamma_{u_{j-1}}^i\big) e^{\theta_j^i}\right),$$

then the laws

$$\mathcal{L}\big(W_N(u_1) - W_N(u_0), \ldots, W_N(u_k) - W_N(u_{k-1}) | \mathcal{N}(1) = N\big)$$

satisfy the large deviations principle with the rate function

$$\begin{aligned}\Lambda^*(\mathbf{z}) &:= \sup_{\boldsymbol{\theta} \in \mathbf{R}^{dk}} \left(\sum_{i,j} \theta_j^i \big(z_{u_j}^i - z_{u_{j-1}}^i\big) - \Lambda(\boldsymbol{\theta})\right) \\ &= \sum_{i,j} \big(z_{u_j}^i - z_{u_{j-1}}^i\big) \log\left(\frac{z_{u_j}^i - z_{u_{j-1}}^i}{\gamma_{u_j}^i - \gamma_{u_{j-1}}^i}\right),\end{aligned}$$

which is precisely the stated form. ◻

We are now ready to give the large deviations principle for the conditional laws on the rescaled paths W_N.

Theorem 2.2. For every $N \geq 0$, let

$$\mu_N = \mathcal{L}\big(W_N | \mathcal{N}(1) = N\big).$$

Then, as a family of laws on $\mathcal{C}^d$, $\{\mu_N\}$ satisfies a large deviations principle with scale N and good rate function

$$\mathcal{I}_W(\boldsymbol{\psi}) = \sum_{i=1}^{d} \int_0^1 I\big(\dot{\psi}^i(t), \dot{\gamma}^i(t)\big) dt$$

when $\boldsymbol{\psi} \in \mathcal{A}_1^d$, and $\mathcal{I}_W(\boldsymbol{\psi}) = \infty$ otherwise.

Proof. By Lemma 2.1, any finite dimensional distribution of difference of the processes $\mathcal{W}_N$ satisfies the large deviations principle. Thus, by [1, Theorem 1], the laws of the processes $\mathcal{W}_N(\cdot)$'s also satisfy the large deviations principle with good rate function

$$\mathcal{I}_W(\boldsymbol{\psi}) = \sum_{i=1}^{d} \int_0^1 I\big(\dot{\psi}^i(u), \dot{\gamma}^i(u)\big) du, \quad \psi^i \in \mathcal{A} \text{ s.t. } \sum_{i=1}^{d} \psi^i(1) = 1,$$

and ∞ otherwise. $\square$

Note that the above large deviations principle are for the processes $\{W_N(\cdot)\}$, where the time parametrisation is random and cannot be observed from the signature. We thus need to parametrise the paths $\mathcal{W}_N$'s at unit speed. For this reason, we introduce the random time change below.

We still condition on $\mathcal{N}(1) = N$. For $j = 1, \dots, N$, let $\tau_j \in [0,1]$ denote the arrival time of the j-th word in the process W, so we have

$$\mathcal{N}(t) = j, \qquad t \in [\tau_j, \tau_{j+1}).$$

Let the random map $T_N : [0,1] \to [0, \tau_N]$ be such that

$$T_N(q) = \tau_j + (Nq - j)(\tau_{j+1} - \tau_j), \qquad q \in \Big(\frac{j}{N}, \frac{j+1}{N}\Big].$$

This says $T_N(j/N)$ is the arrival time of the j-th word, and linearly interpolate in between. Thus, T_N is almost surely a strictly increasing map with inverse $Q_N : [0, \tau_N] \to [0,1]$ defined by

$$Q_N(t) = \frac{j}{N} + \frac{t - \tau_j}{N(\tau_{j+1} - \tau_j)}, \qquad t \in \big(\tau_j, \tau_{j+1}\big].$$

Thus, for every realisation such that $\mathcal{N}(1) = N$, the random path $W_N \circ T_N$ is parametrised at unit speed. The intuition is that the map

$$W_N \circ T_N : [0,1] \to (\mathbf{Z}/N)^d, \qquad q \mapsto W_N(T_N(q))$$

takes the (Nq)-th arrival of the N letters to the position "q" of the path W_N. We let

$$X_N = W_N \circ T_N,$$

which is nothing but the lattice path W_N re-parametrised at unit speed. To investigate the LDP for X_N, note that $T_N = Q_N^{-1}$, and the operations

$$(W_N, Q_N) \mapsto (W_N, T_N) \mapsto W_N \circ T_N$$

are both continuous. Thus, by contraction principle, it suffices to prove the LDP for (W_N, Q_N). We give it in the following lemma.

Lemma 2.3. The laws for $(W_N, Q_N) \in \mathcal{C}^d \times \mathcal{C}$ conditioned on $\mathcal{N}(1) = N$ satisfy the large deviations principle with scale N and good rate function

$$\mathcal{I}_{(W,Q)}(\boldsymbol{\psi}, \phi) = \sum_{i=1}^{d} \int_0^1 I\big(\dot{\psi}^i(t), \dot{\gamma}^i(t)\big) dt$$

for $\boldsymbol{\psi} \in \mathcal{A}_1^d$, $\phi \in \mathcal{A}$ such that $\sum_i \psi^i = \phi$, and $\mathcal{I}_{(W,Q)} = \infty$ otherwise.

Proof. The definition of Q_N ensures that

$$Q_N(t) = \sum_i W_N^i(t)$$

for every $t \in [0,1]$, so the rate function for the pair (W, Q) is the same as $\mathcal{I}_W$ except that one further requires $\sum_i \psi^i = \phi$. Note that this constraint, together with $\boldsymbol{\psi} \in \mathcal{A}_1^d$, forces $\phi(1) = 1$. □

Corollary 2.4. The laws $\mathcal{L}(X_N|\mathcal{N}(1) = N)$ on $\mathcal{C}^d$ satisfies a large deviations principle.

The rate function for X can be expressed in terms of $\mathcal{I}_{(W,Q)}$ using the Contraction Principle [2] but it does not have a closed form. A nice observation from [3] suggests that we can add the component T_N so that the pair (X_N, T_N) satisfies the large deviations principle with a closed form rate function. This is the content of the following theorem.

Theorem 2.5. The conditional laws

$$\nu_N = \mathcal{L}\big((X_N, T_N)|\mathcal{N}(1) = N\big)$$

satisfy a large deviations principle with scale N and good rate function

$$\mathcal{I}_{(X,T)}\big(\boldsymbol{\zeta}, \xi\big) = \mathcal{I}_{(W,Q)}\big(\boldsymbol{\zeta} \circ \xi^{-1}, \xi^{-1}\big) = \sum_{i=1}^{d} \int_0^1 I\big(\dot{\zeta}^i(q), (\gamma^i \circ \xi)'(q)\big) dq \tag{2.3}$$

for $\zeta^i, \xi \in \mathcal{A}$ such that $\sum_i \dot{\zeta}^i \equiv 1$, and $\mathcal{I}_{(X,T)} = \infty$ otherwise.

Proof. Since $T_N = Q_N^{-1}$, it follows directly from the large deviations for inverse processes (see e.g. [3, Theorem 4]) and Lemma 2.3 that

$$\mathcal{I}_{(X,T)}\big(\boldsymbol{\zeta}, \xi\big) = \mathcal{I}_{(W,Q)}\big(\boldsymbol{\zeta} \circ \xi^{-1}, \xi^{-1}\big) = \sum_{i=1}^{d} \int_0^1 I\Big(\big(\zeta^i \circ \xi^{-1}\big)'(t), \dot{\gamma}^i(t)\Big) dt.$$

To derive the specific form of the rate function (the second equality in (2.3)), we note that for each i, we have

$$\big(\zeta^i \circ \xi^{-1}\big)'(t) = \frac{\dot{\zeta}^i(\xi^{-1}(t))}{\dot{\xi}(\xi^{-1}(t))},$$

where for the purpose of the display, we have made an abuse use of notation with both $'$ and $\cdot$ to denote the derivative of a path. A change of variable $q = \xi^{-1}(t)$ then gives

$$\begin{aligned}
\int_0^1 I\Big(\big(\zeta^i \circ \xi^{-1}\big)'(t), \dot{\gamma}^i(t)\Big) dt &= \int_0^1 I\Big(\frac{\dot{\zeta}^i\big(\xi^{-1}(t)\big)}{\dot{\xi}\big(\xi^{-1}(t)\big)}, \dot{\gamma}^i(t)\Big) dt \\
&= \int_0^1 I\Big(\frac{\dot{\zeta}^i(q)}{\dot{\xi}(q)}, \dot{\gamma}^i(\xi(q))\Big) d\xi(q) \\
&= \int_0^1 \dot{\zeta}^i(q) \log\Big(\frac{\dot{\zeta}^i(q)}{\dot{\gamma}^i(\xi(q))\dot{\xi}(q)}\Big) dq.
\end{aligned}$$

The constraint that $\xi \in \mathcal{A}$ is obvious. For the constraint on ζ, we notice that by Lemma 2.3, we need

$$\sum_{i=1}^{d} \zeta^i(\xi^{-1}(q)) = \xi^{-1}(q), \qquad \forall q \in [0,1].$$

This is equivalent as $\sum_i \psi^i(t) = t$ for all t, or $\sum_i \dot{\psi}^i \equiv 1$. □

2.3 Connections with the symmetrisation procedure

In [6], the authors used a procedure of symmetrisation to produce a deterministic sequence of piecewise linear approximations from the signature of a $\mathcal{C}^1$ path to the true path. The construction of that sequence requires rather complicated operations beyond symmetrisation between terms in the signatures. The aim of this subsection is to show that, in the case of monotone paths, these piecewise linear paths can also be "sampled" in a straightforward way after symmetrisation. This turns out to be a simple consequence of the large deviations principle for lattice path sampling.

We first briefly recall the symmetrisation procedure on signatures used in [6], and will mainly follow the notations there. For every integer $N \geq 0$ and $k \geq 0$, let $\mathcal{P}_{N,k}$ denote the set of k-partitions of N; that is,

$$\mathcal{P}_{N,k} = \Big\{ \mathbf{n} = \big(n_1, \ldots, n_k\big) : n_j \geq 0, \ \ \sum_{j=1}^{k} n_j = N \Big\}.$$

For $\mathbf{n} \in \mathcal{P}_{N,k}$, let

$$\mathcal{L}_k^{\mathbf{n}} = \Big\{ \boldsymbol{\ell} = \big(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_k\big) : \boldsymbol{\ell}_j = \big(\ell_j^1, \ldots, \ell_j^d\big), \ \ \sum_{i=1}^{d} \ell_j^i = n_j, \ \ \forall j = 1, \ldots, k \Big\}.$$

Now, for $\mathbf{n} \in \mathcal{P}_{N,k}$ and $\boldsymbol{\ell} \in \mathcal{L}_k^{\mathbf{n}}$, define the set of words $\mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})$ by

$$\mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell}) = \Big\{ w = w_1 * \cdots * w_k : |w_j|_{e_i} = \ell_j^i, \ \ \forall i = 1, \ldots, d, \ \ j = 1, \ldots, k \Big\},$$

where $|w_j|_{e_i}$ denotes the number of the letter e_i in the word w_j. We then define the symmetrised signatures by

$$\mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell}) := N! \sum_{w \in \mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})} C(w).$$

In other words, $\mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})$ is the sum of the coefficients of all words w such that $w = w_1 * \cdots * w_k$, and the number of letters e_i in w_j is ℓ_j^i.

Note that in [6], the symmetrisation procedure is taken with $n_j \equiv n$ for all j, and $N = kn$, so the set-up above is a slight generalisation of that in [6]. Recall the random word W generated by the Poisson process associated to the path γ; we have

$$\mathbf{P}^N\big(W \in \mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})\big) = \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell}).$$

Thus, each $\mathcal{W}_k^{\mathbf{n}}$ corresponds to a random piecewise linear path, which we call $Y_k^{\mathbf{n}}$. We have the following theorem.

Theorem 2.6. For every $N \geq 0$, let $k = k(N)$ and $\mathbf{n} \in \mathcal{P}_{N,k(N)}$ be such that

$$k(N) \to +\infty, \quad \text{and} \quad \sup_{1 \leq j \leq k(N)} \frac{n_j}{N} \to 0$$

as $N \to +\infty$. Then, the sequence $(Y_k^{\mathbf{n}}, T_N)$ satisfies the large deviations principle with the same rate function as in Theorem 2.5.

Proof. It suffices to show that $Y_k^{\mathbf{n}}$ and X_N are exponentially equivalent. In fact, for every realisation of the lattice path X_N, $Y_k^{\mathbf{n}}$ is its polygonal approximation such that its j-th piece connects the points $X_N\big(\bar{n}_{j-1}/N\big)$ and $X_N\big(\bar{n}_j/N\big)$, where

$$\bar{n}_j = \sum_{\ell=1}^{j} n_\ell.$$

Thus, the difference between the j-th piece of $Y_k^{\mathbf{n}}$ and the corresponding part in X_N is at most $\frac{n_j}{N}$, and hence

$$\|Y_k^{\mathbf{n}}(q) - X_N(q))\|_\infty \leq \sup_j \frac{n_j}{N}.$$

Thus, we have

$$\mathbf{P}^N\big(\|Y_k^{\mathbf{n}} - X_N\|_\infty \geq \delta\big) \leq \mathbf{P}^N\Big(\sup_j \frac{n_j}{N} \geq \delta\Big) = 0$$

for all sufficiently large N. This proves the exponential equivalence of $(Y_k^{\mathbf{n}}, T_N)$ and (X_N, T_N), and hence the LDP follows. □

3 A numerical example

In this section, we will provide a numerical example for the sampling scheme (with some modification) introduced above. The example (path) that we are using is the following.

Example 3.1. Let $\gamma : [0,1] \to \mathbf{R}^2$ be the path

$$\gamma(t) = te_1 + t^2 e_2 = (t, t^2), \qquad \forall t \in [0,1].$$

Thus, γ is a two-dimensional monotone path with length 2.

Ideally, in order to test the practical efficiency of the sampling scheme in Section 2.2, one would like to compute the level-N signature of γ and use them as "weights" for lattice paths with N steps. But such a procedure is hard to be visualised when N is large. Thus, for the convenience of simulation and display, instead of trying to sample a lattice path of N steps, we will show the accuracy of the recovery of some of the points in the path.

The procedure we use below is actually a further simplification to the symmetrisation procedure in Section 2.3. Following the notations in Section 2.3, we use $n_j \equiv n$, and we will illustrate reconstructing the path in Example 3.1 in the following two cases:

1. Fix $k = 2$, and vary $n = 3, \ldots, 8$;
2. Fix $n = 4$, and take $k = 2, 3, 4$.

In the first case, we aim to recover the middle point and test the accuracy for different values of n's. For the second one, we fix the length of each block ($n = 4$), and will see that the approximation becomes better when the number of blocks increase.

As for the computation of signature of γ, we compute the truncated signature of the piecewise linear approximation of γ with the time mesh 0.01, and apply the inversion algorithm outlined in the above section. In the first case when we fix $k = 2$, we obtain for each n the weights for the possible $(n+1)^2$ two-piece path[2]. Again, for the simplicity of visualisation, instead of displaying the weights for all $(n+1)^2$ paths, we only draw the path that has the maximum weight for each n (which we call an MLE path).

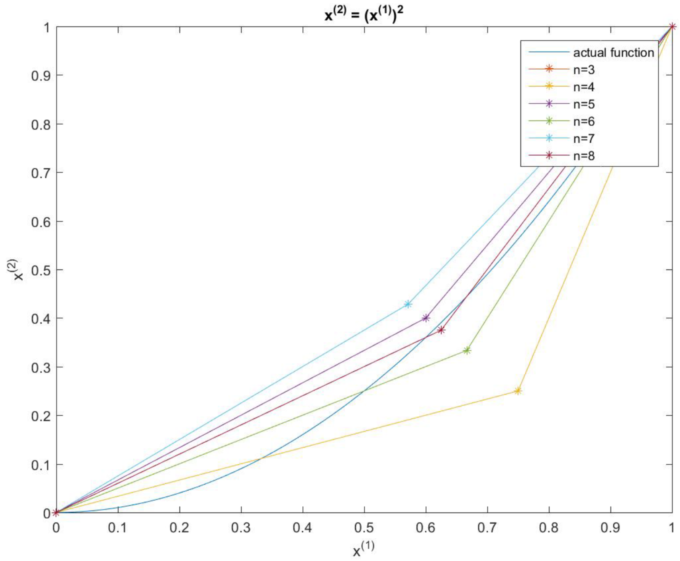

Figure 1: MLE estimators of γ for $k = 2$ and $3 \leq n \leq 8$

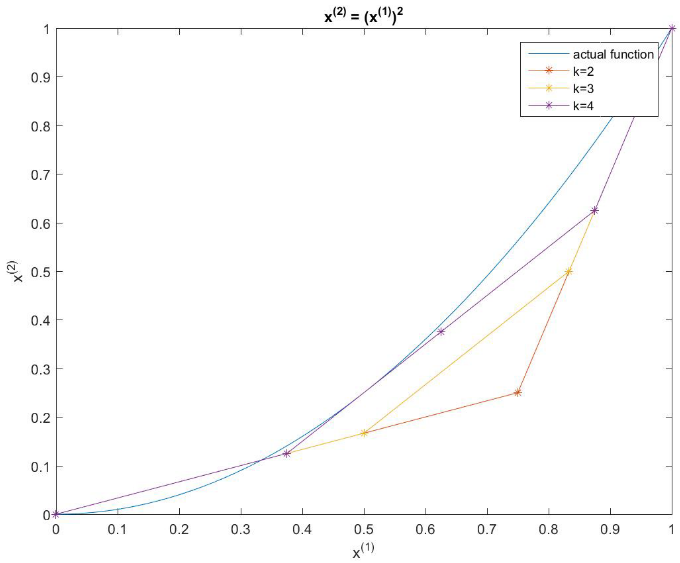

Figure 2: MLE estimators of γ for $n = 4$ and $k = 2, 3, 4$

The two-piece linear paths with the biggest weight for $k = 2$ and $n = 3, \ldots, 8$ are plotted in Figure 1. One sees that the MLE estimator is closer to the true path $\gamma(t) = (t, t^2)$ as n increases. For the second case, we also display the MLE paths for different values of k's in Figure 2, and we see that the approximation is better when k increases.

[2] Recall that the symmetrisation procedure recovers the direction of each segment (or the ratio between the increments in d Euclidean directions). In our two dimensional example, when the block size is n, there are $(n+1)^2$ possible ratios in total, as both the numerator and denominator take value in $\{0, 1, \ldots, n\}$.

References

[1] Dembo, A. and Zajic, T.: Large deviations: From empirical mean and measure to partial sums process. *Stoch. Proc. Appl.*, **57**, (1995), 191–224. MR-1334969

[2] Dembo, A. and Zeitouni, O.: Large Deviations Techniques and Applications. *Springer*, (1998). MR-1619036

[3] Duffy, K. and Rodgers-Lee: Some useful functions for functional large deviations. *Stochastics and Stochastic Reports* **76**, no.3, (2004), 267–279. MR-2072383

[4] Hambly, B. and Lyons, T.: Some notes on trees and paths. arXiv:math.0809.1365.

[5] Hambly, B. and Lyons, T.: Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. Math.*, **171**, no.1, (2010), 109–167. MR-2630037

[6] Lyons, T. and Xu, W.: Inverting the signature of a path. *J. Eur. Math. Soc.*, to appear.

Acknowledgments. WX thanks Terry Lyons for very helpful discussions, especially for telling him the probabilistic interpretation of signatures for monotone paths. We also thank the referee for carefully reading the first manuscript and providing suggestions on improving the presentation.

JC and HN are partially supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1. JC is also supported by ERC Advanced Grant ESig (agreement no. 291244). ND is an Associate Member of the Oxford-Man Institute of Quantitative Finance, which partially supported this work with a visitorship during July 2015. WX is supported by EPSRC through the research fellowship EP/N021568/1.