On Bernstein processes of maximal entropy

Pierre-A Vuillermot

On Bernstein processes of maximal entropy

Pierre-A. Vuillermot*,**
UMR-CNRS 7502, Inst. Élie Cartan de Lorraine, Nancy, France*
Grupo de Física Matemática, GFMUL, Faculdade de Ciências,
Universidade de Lisboa, 1749-016 Lisboa, Portugal**

Abstract

In this article we define and investigate statistical operators and an entropy functional for Bernstein stochastic processes associated with hierarchies of forward-backward systems of decoupled deterministic linear parabolic partial differential equations. The systems under consideration are defined on open bounded domains $D \subset \mathbb{R}^d$ of Euclidean space where $d \in \mathbb{N}^+$ is arbitrary, and are subject to Neumann boundary conditions. We assume that the elliptic part of the parabolic operator in the equations is a self-adjoint Schrödinger operator, bounded from below and with compact resolvent in $L^2(D)$. The statistical operators we consider are then trace-class operators defined from sequences of probabilities associated with the point spectrum of the elliptic part in question, which allow the distinction between pure and mixed processes. We prove in particular that the Bernstein processes of maximal entropy are those for which the associated sequences of probabilities are of Gibbs type. We illustrate our results by considering processes associated with a specific hierarchy of forward-backward heat equations defined in a two-dimensional disk.

1 Introduction and outline

The theory of Bernstein (or reciprocal) processes was launched many years ago in [1] following the seminal contribution put forward in the last section of [13]. At the very end of [13], Schrödinger indeed gave a positive answer to the question whether it is possible to generate a reversible diffusion process from a pair of adjoint, deterministic, linear parabolic partial differential equations whose solutions typically display irreversible behavior. The considerations of [13] were based on entropy arguments, and have had many important ramifications and generalizations over the years up to this day, including connections with Optimal Transport Theory and Stochastic Geometric Mechanics (see, e.g., [2], [4], [9], [10], [15], [21] and the many references therein). On the other hand, a systematic and abstract investigation of continuous time versions of the processes was carried out in [8], according to which it became clear that Bernstein processes may exist without any reference to partial differential equations and may admit as state space any topological space countable at infinity. In spite of that, a great deal of attention has recently been paid to the way that such processes may be generated in Euclidean space of arbitrary dimension from certain particular systems of parabolic partial differential equations, thereby allowing one to recast things within the original framework of [13] with the goal of investigating those processes that are not Markovian (see, e.g., [17]-[19]).

It is our purpose here to continue and deepen our analysis of such processes, and accordingly we shall organize the remaining part of this article in the following way: in Section 2 we introduce a hierarchy of forward-backward systems of decoupled, deterministic, linear parabolic partial differential equations defined on open bounded domains of Euclidean space. Those systems are characterized by the fact that the elliptic part of the parabolic operator is, up to a sign, a self-adjoint Schrödinger operator bounded from below and with compact resolvent in standard L^2-space. The hierarchy comes about by associating with each level of the pure point spectrum of the elliptic part a suitable pair of initial-final data. We then proceed by defining what a Bernstein process is, and show how we can construct from the hierarchy we just alluded to a sequence of such processes that are Markovian. This requires the existence of probability measures of a very specific form which we obtain from the initial-final data and the heat kernel of the given system. In Section 2 we also associate with the spectrum of the elliptic part a sequence of probabilities which eventually allows us to construct non-Markovian processes by means of a suitable averaging procedure, as well as the related statistical operators and the entropy functional which we investigate in detail. Those operators are important in that they allow the classification of the processes as pure or mixed, and we prove in particular that the Bernstein processes of maximal entropy are those for which the probabilities in question are of Gibbs type. In Section 3 we illustrate some of our results by considering Bernstein processes generated by a specific hierarchy of forward-backward heat equations and wandering in a two-dimensional disk, ending up with fairly explicit formulae for the corresponding probabilities and expectation values. Finally, we devote Appendix A to the analysis of statistical operators which are more general than that investigated in Section 2, and Appendix B to stating a general result regarding the very existence of Bernstein processes that goes back to [8] and [18], which we slightly reformulated for the needs of this article. We conclude Appendix B with a brief remark regarding the connection between Bernstein processes, Schrödinger's problem and Optimal Transport Theory.

2 Statistical operators and an entropy functional for Bernstein processes

Let $D \subset \mathbb{R}^d$ with $d \in \mathbb{N}^+$be an open bounded domain with a sufficiently smooth boundary ∂D and let $L^2(D)$ be the standard Hilbert space of all Lebesgue-measurable, square-integrable complex-valued functions on D with respect to Lebesgue measure, whose inner product and induced norm we shall denote by

$(.,.)_2$ and $\|.\|_2$, respectively. Let us consider the differential operator

$$\mathcal{H} = -\frac{1}{2}\Delta_{\mathsf{x}} + V \tag{1}$$

where Δ_{x} stands for Neumann's Laplacian on D and where the following hypothesis holds for the additional term:

(H$_1$) The function $V : D \mapsto \mathbb{R}$ satisfies $V \in L^p(D)$ where

$$p \in \begin{cases} [1, +\infty] & \text{if } d = 1, \\ (1, +\infty] & \text{if } d = 2, \\ \left[\frac{d}{2}, +\infty\right] & \text{if } d \geq 3 \end{cases}$$

and is bounded from below.

Under these conditions it is well known that (1) admits a self-adjoint realization with compact resolvent in $L^2(D)$ and thereby a pure point spectrum $(\lambda_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ such that $\lambda_{\mathsf{m}} \to +\infty$ as $|\mathsf{m}| := \sum_{j=1}^d m_j \to +\infty$, whose corresponding eigenfunctions $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ constitute an orthonormal basis of $L^2(D)$ and are assumed to be real (see, e.g., Chapter VI in [5], particularly Theorem 1.9). For each $\mathsf{m} \in \mathbb{N}^d$ and $T \in (0, +\infty)$ arbitrary, we then introduce the system of adjoint, deterministic, linear parabolic partial differential equations given by

$$\begin{aligned} \partial_t u(\mathsf{x}, t) &= \frac{1}{2}\Delta_{\mathsf{x}} u(\mathsf{x}, t) - V(\mathsf{x})u(\mathsf{x}, t), \quad (\mathsf{x}, t) \in D \times (0, T]\,, \\ u(\mathsf{x}, 0) &= \varphi_{0,\mathsf{m}}(\mathsf{x}), \quad \mathsf{x} \in D, \\ \frac{\partial u(\mathsf{x}, t)}{\partial n(\mathsf{x})} &= 0, \qquad (\mathsf{x}, t) \in \partial D \times (0, T] \end{aligned} \tag{2}$$

and

$$\begin{aligned} -\partial_t v(\mathsf{x}, t) &= \frac{1}{2}\Delta_{\mathsf{x}} v(\mathsf{x}, t) - V(\mathsf{x})v(\mathsf{x}, t), \ (\mathsf{x}, t) \in D \times [0, T)\,, \\ v(\mathsf{x},T) &= \psi_{T,\mathsf{m}}(\mathsf{x}), \quad \mathsf{x} \in D, \\ \frac{\partial v(\mathsf{x}, t)}{\partial n(\mathsf{x})} &= 0, \qquad (\mathsf{x}, t) \in \partial D \times [0, T)\,, \end{aligned} \tag{3}$$

respectively, where $n(\mathsf{x})$ stands for the unit outer normal to ∂D at the point x and where $\varphi_{0,\mathsf{m}}$, $\psi_{T,\mathsf{m}}$ are real-valued functions to be specified below. In this way we are thus considering a hierarchy of problems of the form (2)-(3), that is, an infinite sequence of pairs of such equations where each pair is associated with a level of the spectrum of (1) through the initial-final data. Furthermore, an essential ingredient in the forthcoming considerations will be the heat kernel (or fundamental solution) associated with (2)-(3), which satisfies

$$\begin{cases} g(\mathsf{x},t,\mathsf{y}) = g(\mathsf{y},t,\mathsf{x}), \\ 0 < g(\mathsf{x},t,\mathsf{y}) \leq c_1 t^{-\frac{d}{2}} \exp\left[-c_2 \frac{|\mathsf{x}-\mathsf{y}|^2}{t}\right] \end{cases} \tag{4}$$

for all $\mathsf{x},\mathsf{y} \in \overline{D}$ and every $t \in (0,T]$ for some $c_{1,2} > 0$. It is indeed the knowledge of $\varphi_{0,\mathsf{m}}$, $\psi_{T,\mathsf{m}}$ and (4) that will allow us to construct sequences of Bernstein processes $Z^{\mathsf{m}}_{\tau\in[0,T]}$ wandering in $\overline{D}$. We begin with the following:

Definition 1. *We say the $\overline{D}$-valued process $Z_{\tau\in[0,T]}$ defined on the complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ is a Bernstein process if*

$$\mathbb{E}\left(b(Z_r)\left|\mathcal{F}_s^+ \vee \mathcal{F}_t^-\right.\right) = \mathbb{E}\left(b(Z_r)\left|Z_s, Z_t\right.\right) \tag{5}$$

$\mathbb{P}$-almost everywhere for every bounded Borel measurable function $b : \overline{D} \mapsto \mathbb{C}$, and for all r, s, t satisfying $r \in (s,t) \subset [0,T]$, where $\mathbb{E}(.|.)$ denotes the conditional expectation on $(\Omega,\mathcal{F},\mathbb{P})$. The σ-algebras in (5) are

$$\mathcal{F}_s^+ := \sigma\left\{Z_\tau^{-1}(F) : \tau \leq s,\ F \in \mathcal{B}(\overline{D})\right\}$$

and

$$\mathcal{F}_t^- := \sigma\left\{Z_\tau^{-1}(F) : \tau \geq t,\ F \in \mathcal{B}(\overline{D})\right\},$$

respectively, where $\mathcal{B}(\overline{D})$ stands for the Borel σ-algebra over $\overline{D}$.

The preceding definition is just one out of many equivalent ways of defining a Bernstein process (see, e.g. [8]). It shows that as soon as Z_s and Z_t are known, the behavior of such a process for $\tau \in [s,t]$ is independent of the statistical information available prior to time s and after time t as encoded in $\mathcal{F}_s^+$ and $\mathcal{F}_t^-$, respectively. In fact, a simple probabilistic argument implies that Relation (5) is equivalent to the statement that the σ-algebra

$$\mathcal{F}_{[s,t]} := \sigma\left\{Z_\tau^{-1}(F) : \tau \in [s,t],\ F \in \mathcal{B}(\overline{D})\right\}$$

is conditionally independent of $\mathcal{F}_s^+ \vee \mathcal{F}_t^-$ when

$$\mathcal{F}_{\{s,t\}} := \sigma\left\{Z_s^{-1}(F),\ Z_t^{-1}(F) \ : F \in \mathcal{B}(\overline{D})\right\}$$

is given (see, e.g., Section 25 of Chapter VII in [11] for the notion of conditionally independent σ-algebra). Aside from this property which generalizes Markov's, it is also clear that the above definition maintains a perfect symmetry between past and future in that the σ-algebras $\mathcal{F}_s^+$ and $\mathcal{F}_t^-$ play an identical rôle. Let us now assume that $\varphi_{\mathsf{m},0} > 0$ and $\psi_{\mathsf{m},T} > 0$ are sufficiently smooth on $\overline{D}$ and let us consider the probability measures

$$\mu_{\mathsf{m}}(G) = \int_G \mathsf{dxdy}\varphi_{\mathsf{m},0}(\mathsf{x})g(\mathsf{x},T,\mathsf{y})\psi_{\mathsf{m},T}(\mathsf{y}) \tag{6}$$

for every $G \in \mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$, which satisfy

$$\int_{D\times D} \mathsf{dxdy}\varphi_{\mathsf{m},0}(\mathsf{x})g(\mathsf{x},T,\mathsf{y})\psi_{\mathsf{m},T}(\mathsf{y}) = 1 \tag{7}$$

where g is the heat kernel (4) pinned down at the terminal time T. Then, writing

$$u_{\mathsf{m}}(\mathsf{x},t) = \int_D \mathsf{dy}g(\mathsf{x},t,\mathsf{y})\varphi_{\mathsf{m},0}(\mathsf{y}) > 0 \tag{8}$$

for the solution to (2) and

$$v_{\mathsf{m}}(\mathsf{x},t) = \int_D \mathsf{dy}g(\mathsf{x},T-t,\mathsf{y})\psi_{\mathsf{m},T}(\mathsf{y}) > 0 \tag{9}$$

for the solution to (3), we have the following result which follows from the substitution of (6) into the formulae of Theorem B.1 of Appendix B, and from Theorem 2 in [18] as far as the Markov property is concerned:

Theorem 1. *Assume that Hypothesis* (H_1) *holds. Then for every* $\mathsf{m} \in \mathbb{N}^d$ *there exists a probability space* $\left(\Omega, \mathcal{F}, \mathbb{P}_{\mu_{\mathsf{m}}}\right)$ *supporting a* $\overline{D}$*-valued Bernstein process* $Z^{\mathsf{m}}_{\tau\in[0,T]}$ *such that the following statements are valid:*

(a) The process $Z^{\mathsf{m}}_{\tau\in[0,T]}$ *is a forward Markov process whose finite-dimensional distributions are*

$$\begin{aligned} &\mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_{t_1} \in F_1, ..., Z^{\mathsf{m}}_{t_n} \in F_n\right) \\ = &\int_D \mathsf{dx}\rho_{\mathsf{m},0}(\mathsf{x}) \int_{F_1} \mathsf{dx}_1 ... \int_{F_n} \mathsf{dx}_n \prod_{k=1}^{n} w^*_{\mathsf{m}}\left(\mathsf{x}_{k-1}, t_{k-1}; \mathsf{x}_k, t_k\right) \end{aligned} \tag{10}$$

for every $n \in \mathbb{N}^+$, *all* $F_1, ..., F_n \in \mathcal{B}(\overline{D})$ *and all* $0 = t_0 < t_1 < ... < t_n < T$, *with* $\mathsf{x}_0 = \mathsf{x}$. *In the preceding expression the density of the forward Markov transition function is*

$$w^*_{\mathsf{m}}(\mathsf{x}, s; \mathsf{y}, t) = g(\mathsf{x}, t-s, \mathsf{y}) \frac{v_{\mathsf{m}}(\mathsf{y}, t)}{v_{\mathsf{m}}(\mathsf{x}, s)} \tag{11}$$

for all $\mathsf{x}, \mathsf{y} \in \overline{D}$ *and all* $s, t \in [0,T]$ *with* $t > s$, *while the initial distribution of the process reads*

$$\rho_{\mathsf{m},0}(\mathsf{x}) = \varphi_{\mathsf{m},0}(\mathsf{x}) v_{\mathsf{m}}(\mathsf{x}, 0). \tag{12}$$

(b) The process $Z^{\mathsf{m}}_{\tau\in[0,T]}$ *may also be viewed as a backward Markov process since the finite-dimensional distributions (10) may also be written as*

$$\begin{aligned} &\mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_{t_1} \in F_1, ..., Z^{\mathsf{m}}_{t_n} \in F_n\right) \\ = &\int_D \mathsf{dx}\rho_{\mathsf{m},T}(\mathsf{x}) \int_{F_1} \mathsf{dx}_1 ... \int_{F_n} \mathsf{dx}_n \prod_{k=1}^{n} w_{\mathsf{m}}\left(\mathsf{x}_{k+1}, t_{k+1}; \mathsf{x}_k, t_k\right) \end{aligned} \tag{13}$$

for every $n \in \mathbb{N}^{+}$*, all* $F_1, ..., F_n \in \mathcal{B}(\overline{D})$ *and all* $0 < t_1 < ... < t_n < t_{n+1} = T$*, with* $\mathsf{x}_{n+1} = \mathsf{x}$*. In the preceding expression the density of the backward Markov transition function is*

$$w_{\mathsf{m}}(\mathsf{x}, t; \mathsf{y}, s) = g(\mathsf{x}, t - s, \mathsf{y}) \frac{u_{\mathsf{m}}(\mathsf{y}, s)}{u_{\mathsf{m}}(\mathsf{x}, t)} \tag{14}$$

for all $\mathsf{x}, \mathsf{y} \in \overline{D}$ *and all* $s, t \in [0, T]$ *with* $t > s$*, while the final distribution of the process reads*

$$\rho_{\mathsf{m},T}(\mathsf{x}) = \psi_{\mathsf{m},T}(\mathsf{x}) u_{\mathsf{m}}(\mathsf{x}, T).$$

(c) We have

$$\mathbb{P}_{\mu_{\mathsf{m}}}\left(Z_t^{\mathsf{m}} \in F\right) = \int_F \mathsf{d}\mathsf{x}\, u_{\mathsf{m}}(\mathsf{x}, t) v_{\mathsf{m}}(\mathsf{x}, t) \tag{15}$$

for each $t \in [0, T]$ *and every* $F \in \mathcal{B}(\overline{D})$*.*

(d) Finally,

$$\mathbb{E}_{\mu_{\mathsf{m}}}\left(b(Z_t^{\mathsf{m}})\right) = \int_D \mathsf{d}\mathsf{x}\, b(\mathsf{x}) u_{\mathsf{m}}(\mathsf{x}, t) v_{\mathsf{m}}(\mathsf{x}, t) \tag{16}$$

for each bounded Borel measurable function $b : \overline{D} \mapsto \mathbb{C}$ *and every* $t \in [0, T]$*.*

Remarks. (1) The fact that $Z^{\mathsf{m}}_{\tau \in [0,T]}$ is a Markov process for each m may be read off Relations (10) and (13), inasmuch as (11) and (14) are the densities of transition functions that satisfy the Chapman-Kolmogorov equation (see, e.g., Lemmas 1 and 2 in [18], and for more general comments Section 2.4 in Chapter 2 of [7]). Alternatively, the Markov property of $Z^{\mathsf{m}}_{\tau \in [0,T]}$ is an immediate consequence of the form (6) of the underlying probability measures through Theorem 3.1 in [8]. Furthermore, the fact that $Z^{\mathsf{m}}_{\tau \in [0,T]}$ is both a forward and a backward Markov process is related to the perfect symmetry between past and future which we alluded to above, also encoded in (15) where (8) and (9) play an equivalent rôle. We refer the reader to [18] for further considerations on this issue, where a general notion of reversibility was put forward in order to deal with processes generated by systems of non-autonomous forward-backward parabolic equations. Finally, we note that the processes $Z^{\mathsf{m}}_{\tau \in [0,T]}$ are in general non-stationary (see, e.g., our construction in Section 3).

(2) Theorem 3.1 in [8] actually says much more than what we just referred to in the preceding remark. Indeed, when applied to the present situation, it asserts that one may generate a Markovian Bernstein process from a probability measure μ on $\mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$ if, and only if, there exist positive measures ν_0 and ν_T on $\mathcal{B}(\overline{D})$ such that

$$\mu\left(G\right) = \int_G \mathsf{d}\left(\nu_0 \otimes \nu_T\right)(\mathsf{x}, \mathsf{y})\, g(\mathsf{x}, T, \mathsf{y}) \tag{17}$$

for every $G \in \mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$, with $\mu\left(D \times D\right) = 1$. It provides therefore a very simple and practical criterion to decide whether a Bernstein process is Markovian or not.

It is consequently easy to generate non-Markovian processes out of those constructed in Theorem 1. One possible way to achieve that and eventually define the statistical operators and the entropy functional we are interested in amounts to associating a sequence $(p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ of probabilities with the pure point spectrum of (1), that is, a sequence of numbers satisfying

$$p_{\mathsf{m}} \geq 0, \qquad \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} = 1, \tag{18}$$

and to consider weighted averages of the form

$$\overline{\mu}(G) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\mu_{\mathsf{m}}(G) \tag{19}$$

where $\mu_{\mathsf{m}}(G)$ is given by (6). We note that the preceding series is indeed convergent and defines a genuine probability measure by virtue of (7) and (18). However, in order to generate non-Markovian processes from (19) we ought to identify its joint probability density in view of Remark 2. To this end and aside from having the smoothness of $\varphi_{\mathsf{m},0} > 0$ and $\psi_{\mathsf{m},T} > 0$ on $\overline{D}$, the following additional hypothesis turns out to be sufficient:

(H_2) We have

$$\sup_{\mathsf{m}\in\mathbb{N}^d}\sup_{\mathsf{x}\in\overline{D}} \varphi_{\mathsf{m},0}(\mathsf{x}) < +\infty$$

and

$$\sup_{\mathsf{m}\in\mathbb{N}^d}\sup_{\mathsf{x}\in\overline{D}} \psi_{\mathsf{m},T}(\mathsf{x}) < +\infty.$$

This hypothesis indeed clearly implies that

$$\sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\varphi_{\mathsf{m},0}(\mathsf{x})\,\psi_{\mathsf{m},T}(\mathsf{y}) < +\infty \tag{20}$$

for all $\mathsf{x},\mathsf{y} \in \overline{D}$, so that the joint probability density associated with (19) may be written as

$$\overline{\mu}(\mathsf{x},\mathsf{y}) = g(\mathsf{x},T,\mathsf{y}) \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\varphi_{\mathsf{m},0}(\mathsf{x})\psi_{\mathsf{m},T}(\mathsf{y}). \tag{21}$$

Then the following result holds:

Theorem 2. *Assume that Hypotheses* (H_1) *and* (H_2) *hold, and for every* $\mathsf{m} \in \mathbb{N}^d$ *let* $Z^{\mathsf{m}}_{\tau\in[0,T]}$ *be the process of Theorem 1. Let* $\overline{Z}_{\tau\in[0,T]}$ *be the Bernstein process obtained by substituting (19) into the formulae of Theorem B.1. Then the following statements are valid:*

(a) The finite-dimensional distributions of the process $\overline{Z}_{\tau\in[0,T]}$ *are*

$$\begin{aligned} &\mathbb{P}_{\overline{\mu}}\left(\overline{Z}_{t_1} \in F_1, ..., \overline{Z}_{t_n} \in F_n\right) \\ = &\sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_{t_1} \in F_1, ..., Z^{\mathsf{m}}_{t_n} \in F_n\right) \end{aligned}$$

for every $n \in \mathbb{N}^+$ and all $F_1, ..., F_n \in \mathcal{B}(\overline{D})$, where $P_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_{t_1} \in F_1, ..., Z^{\mathsf{m}}_{t_n} \in F_n\right)$ is given either by (10) or (13). In addition, if (20) is not of the form $\nu_0 \otimes \nu_T$ where ν_0 and ν_T are as in (17) then $\overline{Z}_{\tau\in[0,T]}$ is non-Markovian.

(b) We have

$$\mathbb{P}_{\overline{\mu}}\left(\overline{Z}_t \in F\right) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_t \in F\right) \tag{22}$$

for each $t \in [0, T]$ and every $F \in \mathcal{B}(\overline{D})$, where $P_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_t \in F\right)$ is given by (15).

(c) We have

$$\mathbb{E}_{\overline{\mu}}\left(b(\overline{Z}_t)\right) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \mathbb{E}_{\mu_{\mathsf{m}}}\left(b(Z^{\mathsf{m}}_t)\right) \tag{23}$$

for each bounded Borel measurable function $b : \overline{D} \mapsto \mathbb{C}$ and every $t \in [0, T]$, where $E_{\mu_{\mathsf{m}}}\left(b(Z^{\mathsf{m}}_t)\right)$ is given by (16).

Proof. It follows from Theorem B.1 of the Appendix that a Bernstein process generated from a statistical mixture of probability measures coincides with the statistical mixture of the processes generated from those measures, so that Theorem 2 follows immediately from Theorem 1 and (19). The fact that the process $\overline{Z}_{\tau\in[0,T]}$ is non-Markovian when the structural hypothesis regarding (20) holds is a direct consequence of Remark 2. ■

Remark. The structural hypothesis we just referred to is necessary in that it allows one to disregard cases like $\varphi_{\mathsf{m},0} = \varphi_0$ or $\psi_{\mathsf{m},T} = \psi_T$ for every m, or the situation where $p_{\mathsf{m}^*} = 1$ for some $\mathsf{m}^* \in \mathbb{N}^d$, among others. Indeed, initial or final data that are identical for each level of the spectrum still lead to a joint density like that of (17) with $\nu_0 = \varphi_0$ or $\nu_T = \psi_T$, and hence to a Markovian process as is the case when $p_{\mathsf{m}^*} = 1$ for some $\mathsf{m}^* \in \mathbb{N}^d$. We shall dwell a bit more on this further below when we deal with the example in Section 3.

We now enquire about the possibility of choosing

$$\begin{cases} \varphi_{\mathsf{m},0} = \mathsf{f}_{\mathsf{m}}, \\ \psi_{\mathsf{m},T} = \exp\left[T\lambda_{\mathsf{m}}\right] \mathsf{f}_{\mathsf{m}} \end{cases} \tag{24}$$

as initial-final-data in (8) and (9), where $(\lambda_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ and $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ stand for the eigenvalues and eigenfunctions of (1), respectively. The difficulty is that the eigenfunctions f_{m} are not positive in general with the possible exception of f_0, so that the μ_{m} are no longer positive measures with the possible exception of μ_{0}. Therefore, we may not associate a Bernstein process with each level of the spectrum as we did in Theorem 1. Nevertheless, we proceed by showing that the above averaging method still allows us to get genuine probability measures in certain cases. We begin by proving that the μ_{m} satisfy the correct normalization condition under an additional hypothesis:

Lemma 1. *For each* $\mathsf{m} \in \mathbb{N}^d$*, let us consider measures* μ_{m} *of the form (6) where* $\varphi_{\mathsf{m},0}$ *and* $\psi_{\mathsf{m},T}$ *are given by (24). Then* μ_{m} *is a signed measure. Moreover, if*

$$\mathcal{Z}(T) := \sum_{\mathsf{n}\in\mathbb{N}^d} \exp\left[-T\lambda_{\mathsf{n}}\right] < +\infty \tag{25}$$

we have

$$\mu_{\mathsf{m}}\left(D \times D\right) = 1. \tag{26}$$

Proof. We have just explained why μ_{m} is not a positive measure, so that we need only prove (26). Since (25) holds we have the spectral decomposition

$$g(\mathsf{x}, T, \mathsf{y}) = \sum_{\mathsf{n}\in\mathbb{N}^d} \exp\left[-T\lambda_{\mathsf{n}}\right] \mathsf{f}_{\mathsf{n}}(\mathsf{x})\mathsf{f}_{\mathsf{n}}(\mathsf{y}) \tag{27}$$

as a strongly convergent series in $L^2(D \times D)$ for heat kernel (4). Therefore, from (6) and (24) we obtain

$$\begin{aligned}\mu_{\mathsf{m}}(D \times D) &= \exp\left[T\lambda_{\mathsf{m}}\right] \int_{D\times D} \mathsf{dxdy}\mathsf{f}_{\mathsf{m}}(\mathsf{x}) g(\mathsf{x}, T, \mathsf{y}) \mathsf{f}_{\mathsf{m}}(\mathsf{y}) \\ &= \sum_{\mathsf{n}\in\mathbb{N}^d} \exp\left[T(\lambda_{\mathsf{m}} - \lambda_{\mathsf{n}})\right] (\mathsf{f}_{\mathsf{m}}, \mathsf{f}_{\mathsf{n}})_2^2 = 1\end{aligned}$$

as a consequence of the orthogonality properties of $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$. ■

Sequences of Gibbs probabilities of the form

$$p_{\mathsf{m}} = \mathcal{Z}^{-1}(T) \exp\left[-T\lambda_{\mathsf{m}}\right] \tag{28}$$

will play an important rôle in the sequel. In fact, with (28) the joint probability density of the statistical mixture of the μ_{m} in Lemma 1 reads

$$\begin{aligned}\overline{\mu}(\mathsf{x}, \mathsf{y}) &= g(\mathsf{x}, T, \mathsf{y}) \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \exp\left[T\lambda_{\mathsf{m}}\right] \mathsf{f}_{\mathsf{m}}(\mathsf{x})\mathsf{f}_{\mathsf{m}}(\mathsf{y}) \\ &= \mathcal{Z}^{-1}(T) g(\mathsf{x}, T, \mathsf{y}) \sum_{\mathsf{m}\in\mathbb{N}^d} \mathsf{f}_{\mathsf{m}}(\mathsf{x})\mathsf{f}_{\mathsf{m}}(\mathsf{y}) \\ &= \mathcal{Z}^{-1}(T) g(\mathsf{x}, T, \mathsf{y}) \delta(\mathsf{x} - \mathsf{y}) \end{aligned} \tag{29}$$

as a consequence of the completeness of the basis $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$. Thus, having (4) and (29) at our disposal, the latter obviously not being of the form (17), and substituting (29) into Theorem B.1 of Appendix B we obtain:

Theorem 3. *Let us assume that Hypothesis* (H$_1$) *holds, and let* $\bar{Z}_{\tau\in[0,T]}$ *be the Bernstein generated by (29). Then the following statements are valid:*

(a) The process $\bar{Z}_{\tau\in[0,T]}$ is stationary, non-Markovian and for every $n \in \mathbb{N}^+$ with $n \geq 2$ its finite-dimensional distributions are

$$\begin{aligned}
& \mathbb{P}_{\bar{\mu}}\left(\bar{Z}_{t_1} \in F_1, ..., \bar{Z}_{t_n} \in F_n\right) \\
= \; & \mathcal{Z}^{-1}(T) \int_{F_1} \mathsf{d}\mathsf{x}_1 ... \int_{F_n} \mathsf{d}\mathsf{x}_n \\
& \times \prod_{k=2}^{n} g\left(\mathsf{x}_k, t_k - t_{k-1}, \mathsf{x}_{k-1}\right) \times g\left(\mathsf{x}_1, T-(t_n - t_1), \mathsf{x}_n\right) \qquad (30)
\end{aligned}$$

for all $F_1, ..., F_n \in \mathcal{B}(\overline{D})$ and all $0 < t_1 < ... < t_n < T$.

(b) We have

$$\mathbb{P}_{\bar{\mu}}\left(\bar{Z}_t \in F\right) = \mathcal{Z}^{-1}(T) \int_F \mathsf{d}\mathsf{x} g\left(\mathsf{x}, T, \mathsf{x}\right) \qquad (31)$$

for each $F \in \mathcal{B}(\overline{D})$ and every $t \in [0, T]$.

(c) We have

$$\mathbb{E}_{\bar{\mu}}\left(b(\bar{Z}_t)\right) = \mathcal{Z}^{-1}(T) \int_D \mathsf{d}\mathsf{x} b(\mathsf{x}) g\left(\mathsf{x}, T, \mathsf{x}\right) \qquad (32)$$

for each bounded Borel measurable function $b : \overline{D} \mapsto \mathbb{C}$ and every $t \in [0, T]$.

Remark. The fact that the process of the preceding result is stationary is tied up with the structure of the finite-dimensional distributions (30), which differs from those in Theorems 1 and 2. Indeed, for any $\tau > 0$ sufficiently small such that $0 < t_1 + \tau < ... < t_n + \tau < T$ we have

$$\mathbb{P}_{\bar{\mu}}\left(\bar{Z}_{t_1+\tau} \in F_1, ..., \bar{Z}_{t_n+\tau} \in F_n\right) = \mathbb{P}_{\bar{\mu}}\left(\bar{Z}_{t_1} \in F_1, ..., \bar{Z}_{t_n} \in F_n\right),$$

as well as the time independence of (31) and (32). Furthermore we also note that since $\mathbb{P}_{\bar{\mu}}\left(\bar{Z}_t \in D\right) = 1$, Relation (31) provides yet another expression for (25), namely

$$\mathcal{Z}(T) = \int_D \mathsf{d}\mathsf{x} g\left(\mathsf{x}, T, \mathsf{x}\right)$$

which, of course, also follows from (27) and the fact that $\left\|\mathsf{f}_{\mathsf{m}}\right\|_2 = 1$ for every $\mathsf{m} \in \mathbb{N}^d$.

The preceding results thus reveal the possibility of having at least two types of Bernstein processes, namely, on the one hand Markovian processes associated with each level of the spectrum of (1), and on the other hand typically non-Markovian processes obtained by averaging $Z^{\mathsf{m}}_{\tau\in[0,T]}$ over the whole spectrum for a given sequence $(p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$, or by averaging signed measures. In order to better characterize those processes by means of entropy considerations, we now proceed by introducing a statistical operator and an entropy functional by analogy with Quantum Statistical Mechanics. We define

$$\mathcal{R}f := \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \left(f, \mathsf{f}_{\mathsf{m}}\right)_2 \mathsf{f}_{\mathsf{m}} \tag{33}$$

for each $f \in L^2(D)$. The following result is elementary, so that we only sketch the proof of the trace-class property which will reappear in Appendix A:

Proposition 1. *Let us assume that Hypothesis* (H_1) *holds. Then the following statements are valid:*

(a) Expression (33) defines a self-adjoint, positive trace-class operator in $L^2(D)$ *such that the inequalities*

$$0 \leq \mathcal{R}^2 \leq \mathcal{R} \leq \mathbb{I}$$

hold in the sense of quadratic forms, where $\mathbb{I}$ *stands for the identity in* $L^2(D)$*. More specifically we have*

$$\mathrm{Tr}\,\mathcal{R} = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} = 1 \tag{34}$$

and

$$\mathrm{Tr}\,\mathcal{R}^2 = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}^2 \leq 1. \tag{35}$$

In particular we have

$$\mathrm{Tr}\,\mathcal{R}^2 = 1$$

if, and only if, $p_{\mathsf{m}^*} = 1$ *for some* $\mathsf{m}^* \in \mathbb{N}^d$ *and thus* $p_{\mathsf{m}} = 0$ *for every* $\mathsf{m} \neq \mathsf{m}^*$.

(b) The eigenvalue equation

$$\mathcal{R}\mathsf{f}_{\mathsf{m}} = p_{\mathsf{m}}\mathsf{f}_m \tag{36}$$

holds for each $\mathsf{m} \in \mathbb{N}^d$ *and the spectrum of* $\mathcal{R}$ *is either pure point with* $\sigma(\mathcal{R}) = (p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ *if* $p_{\mathsf{m}} = 0$ *for at least one* m, *or* $\sigma(\mathcal{R}) = (p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d} \cup \{0\}$ *if* $0 < p_{\mathsf{m}} < 1$ *for every* m, *in which case zero is not an eigenvalue.*

(c) If B *is a linear bounded self-adjoint operator on* $L^2(D)$ *we have*

$$\mathrm{Tr}\left(\mathcal{R}B\right) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \left(B\mathsf{f}_{\mathsf{m}}, \mathsf{f}_{\mathsf{m}}\right)_2. \tag{37}$$

In particular, if B *is the multiplication operator given by* $Bf = bf$ *where* $b \in L^{\infty}(D)$ *is real-valued, we have*

$$\mathrm{Tr}\left(\mathcal{R}B\right) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \int_D d\mathsf{x} b(\mathsf{x}) \left|\mathsf{f}_{\mathsf{m}}(\mathsf{x})\right|^2. \tag{38}$$

Proof. Owing to the properties of p_{m} and f_{m} it is immediate that (33) defines a linear bounded operator in $L^2(D)$. Now let $(\mathsf{h}_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}^d}$ be an arbitrary orthonormal basis in $L^2(D)$. In order to prove that $\mathcal{R}$ is trace-class, it is then necessary and sufficient to show that

$$\sum_{\mathsf{n}\in\mathbb{N}^d} (\mathcal{R}\mathsf{h}_\mathsf{n}, \mathsf{h}_\mathsf{n})_2 < +\infty$$

(see, e.g., Theorem 8.1 in Chapter III of [6]). To this end let us introduce the function

$$a(\mathsf{m},\mathsf{n}) := p_\mathsf{m} \, (\mathsf{h}_\mathsf{n}, \mathsf{f}_\mathsf{m})_2 \, (\mathsf{f}_\mathsf{m}, \mathsf{h}_\mathsf{n})_2 \tag{39}$$

so that

$$\sum_{\mathsf{m}\in\mathbb{N}^d} a(\mathsf{m},\mathsf{n}) = (\mathcal{R}\mathsf{h}_\mathsf{n}, \mathsf{h}_\mathsf{n})_2 \tag{40}$$

for every fixed n. Moreover, for any fixed m we have

$$\sum_{\mathsf{n}\in\mathbb{N}^d} a(\mathsf{m},\mathsf{n}) = p_\mathsf{m} \tag{41}$$

since $\|\mathsf{f}_\mathsf{m}\|_2 = 1$. Furthermore, the preceding series converges absolutely since from (39) we have for any choice of positive integers $N_1, ..., N_d$ the estimate

$$\begin{aligned} &\sum_{\mathsf{n}:0\leq \mathsf{n}_j \leq N_j} |a(\mathsf{m},\mathsf{n})| \\ \leq \quad & p_\mathsf{m} \left(\sum_{\mathsf{n}\in\mathbb{N}^d} |(\mathsf{h}_\mathsf{n}, \mathsf{f}_\mathsf{m})_2|^2 \right)^{\frac{1}{2}} \left(\sum_{\mathsf{n}\in\mathbb{N}^d} |(\mathsf{f}_\mathsf{m}, \mathsf{h}_\mathsf{n})_2|^2 \right)^{\frac{1}{2}} = p_\mathsf{m} \end{aligned}$$

for any fixed m. Consequently we have

$$\sum_{\mathsf{m}\in\mathbb{N}^d} \sum_{\mathsf{n}\in\mathbb{N}^d} |a(\mathsf{m},\mathsf{n})| \leq \sum_{\mathsf{m}\in\mathbb{N}^d} p_\mathsf{m} = 1,$$

from which we infer according to well-known criterias that the associated iterated series are equal, that is,

$$\sum_{\mathsf{n}\in\mathbb{N}^d} \sum_{\mathsf{m}\in\mathbb{N}^d} a(\mathsf{m},\mathsf{n}) = \sum_{\mathsf{m}\in\mathbb{N}^d} \sum_{\mathsf{n}\in\mathbb{N}^d} a(\mathsf{m},\mathsf{n}).$$

Equivalently, this means that

$$\operatorname{Tr} \mathcal{R} := \sum_{\mathsf{n}\in\mathbb{N}^d} (\mathcal{R}\mathsf{h}_\mathsf{n}, \mathsf{h}_\mathsf{n})_2 = \sum_{\mathsf{m}\in\mathbb{N}^d} p_\mathsf{m} = 1$$

according to (40) and (41), which proves (34). The proof of (35) is similar with

$$\mathcal{R}^2 f = \sum_{\mathsf{m}\in\mathbb{N}^d} p_\mathsf{m}^2 \, (f, \mathsf{f}_\mathsf{m})_2 \, \mathsf{f}_\mathsf{m}.$$

The remaining statements are immediate from elementary arguments. ■

REMARK. Regarding expression (38) we note that when the p_{m} are given by (28) we have

$$\mathrm{Tr}\,(\mathcal{R}B) = \mathbb{E}_{\bar{\mu}}\left(b(\bar{Z}_t)\right) \tag{42}$$

for every $t \in [0,T]$, where the right-hand side is given by (32). This is an immediate consequence of (27), so that the statistical average (38) calculated by means of Gibbs probabilities coincides with the expectation of some function of the process of Theorem 3. This is of course only possible because that process is stationary, the right-hand side of (42) then being time-independent as the left-hand side is. It is therefore reasonable to ask whether relations such as (42) may exist in more general cases, for instance for the averaged processes of Theorem 2 which are in general non-stationary. This is indeed possible as we shall show in the appendix, provided we have at our disposal a class of time-dependent statistical operators which generalize (33).

By analogy with Quantum Statistical Mechanics from which we also borrow the terminology (see, e.g., Section 3 in Chapter V of [16]), Proposition 1 allows us to establish a preliminary classification of the Bernstein processes constructed above, according to the following:

Definition 2. *For a given sequence $(p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ let $\overline{Z}_{\tau\in[0,T]}$ be the Bernstein process of Theorem 2, and let $\mathcal{R}$ be the statistical operator given by (33). If $\mathrm{Tr}\,\mathcal{R}^2 = 1$ we say that $\overline{Z}_{\tau\in[0,T]}$ is a pure process, whereas if $\mathrm{Tr}\,\mathcal{R}^2 < 1$ we say that $\overline{Z}_{\tau\in[0,T]}$ is a mixed process.*

We note that in the first case we necessarily have $\overline{Z}_{\tau\in[0,T]} = Z^{\mathsf{m}^*}_{\tau\in[0,T]}$ for some $\mathsf{m}^* \in \mathbb{N}^d$ according to the second part of (a) in Proposition 1, so that $\overline{Z}_{\tau\in[0,T]}$ reduces to a Markovian process according to Theorem 1 or the remark following the proof of Theorem 2. On the other hand, an important example which illustrates the second case is that of Gibbs probability measures (28).

We now introduce the entropy functional

$$\mathsf{S} := \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \ln\left(\frac{1}{p_{\mathsf{m}}}\right) \tag{43}$$

where we define $x\ln\left(\frac{1}{x}\right)$ to be zero at $x=0$ so that $\mathsf{S} = 0$ if, and only if, $p_{\mathsf{m}} = 0$ or $p_{\mathsf{m}} = 1$ for every m, the latter value being associated with pure processes according to Definition 2. It is plain that we may have $\mathsf{S} = +\infty$ despite the normalization (18), a case in point being that of the Gibbs probabilities (28). Indeed, the substitution of (28) into (43) shows that $\mathsf{S} < +\infty$ if, and only if, the additional condition

$$\sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-T\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}} < +\infty$$

holds. From now on we shall therefore assume that the p_{m} are chosen in such a way that $0 < p_{\mathsf{m}} < 1$ with

$$\sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \ln\left(\frac{1}{p_{\mathsf{m}}}\right) < +\infty. \tag{44}$$

The following result is then our desired optimization statement for (43). Generally speaking the proof of such results requires the Fréchet differentiability of the functionals involved in some appropriate space (see, e.g., Chapter IV in [14]). However, the simple structure of (43) allows us to bypass that requirement, so that we shall provide an independent proof of our theorem for the sake of completeness. In addition we note that we only consider probabilities which assign an *a priori* prescribed value to the average of the spectrum of (1):

Theorem 4. *Let us consider the set of all sequences* $(p_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ *satisfying* $0 < p_{\mathsf{m}} < 1$ *for every* m, *along with*

$$\sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} = 1 \tag{45}$$

and (44). Moreover, let $\lambda \in \mathbb{R}$ *be given and let us assume that*

$$\sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\lambda_{\mathsf{m}} = \lambda. \tag{46}$$

Then the following statements are valid:

(a) There exists a finite constant $\beta(\lambda) > 0$ *such that*

$$\mathcal{Z}(\beta) := \sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-\beta\lambda_{\mathsf{m}}\right] < +\infty \tag{47}$$

and

$$\mathcal{Z}^{-1}(\beta) \sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-\beta\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}} < +\infty \tag{48}$$

for every $\beta \in [\beta(\lambda), +\infty)$.

(b) Among all the mixed processes obtained from sequences of the above type by the method of Theorem 2, the process of maximal entropy is that generated from probabilities given by

$$p_{\mathsf{m}} = \mathcal{Z}^{-1}(\beta(\lambda)) \exp\left[-\beta(\lambda)\lambda_{\mathsf{m}}\right] \tag{49}$$

for every $\mathsf{m} \in \mathbb{N}^d$. *Moreover we have*

$$\mathcal{Z}^{-1}(\beta(\lambda)) \sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-\beta(\lambda)\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}} = \lambda. \tag{50}$$

(c) If we assume in addition that $\sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-\beta\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}} < +\infty$ *for every* $\beta \in (0, \beta(\lambda))$, *then* $\beta \mapsto \mathcal{Z}(\beta)$ *is differentiable at* $\beta = \beta(\lambda)$ *and we have*

$$\mathsf{S}_{\max}(\lambda) = \ln \mathcal{Z}(\beta(\lambda)) - \beta(\lambda)\frac{d}{d\beta}\ln \mathcal{Z}(\beta)_{|\beta=\beta(\lambda)} \tag{51}$$

for the maximal entropy of part (b).

Proof. Since $\lambda_{\mathsf{m}} \to +\infty$ as $|\mathsf{m}| \to +\infty$ and since (45) holds, there exist $\mathsf{n}, \mathsf{n}' \in \mathbb{N}^d$ with $\mathsf{n} \neq \mathsf{n}'$ such that $\lambda_{\mathsf{n}} \neq \lambda_{\mathsf{n}'}$ and $p_{\mathsf{n}} \neq p_{\mathsf{n}'}$. We then consider the inhomogeneous system

$$\alpha + \beta\lambda_{\mathsf{n}} = -(\ln p_{\mathsf{n}} + 1), \tag{52}$$

$$\alpha + \beta\lambda_{\mathsf{n}'} = -(\ln p_{\mathsf{n}'} + 1) \tag{53}$$

in the two unknowns α and β, whose unique solution pair reads

$$\alpha = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} (\lambda_{\mathsf{n}} (\ln p_{\mathsf{n}'} + 1) - \lambda_{\mathsf{n}'} (\ln p_{\mathsf{n}} + 1)), \tag{54}$$

$$\beta = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} \ln \frac{p_{\mathsf{n}}}{p_{\mathsf{n}'}}. \tag{55}$$

Furthermore, let us write (45) and (46) as

$$p_{\mathsf{n}} + p_{\mathsf{n}'} = 1 - \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}},$$

$$p_{\mathsf{n}}\lambda_{\mathsf{n}} + p_{\mathsf{n}'}\lambda_{\mathsf{n}'} = \lambda - \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}}\lambda_{\mathsf{m}},$$

respectively, which gives

$$p_{\mathsf{n}} = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} \left(\lambda_{\mathsf{n}'} \left(1 - \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}} \right) - \lambda + \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}}\lambda_{\mathsf{m}} \right)$$

and

$$p_{\mathsf{n}'} = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} \left(\lambda - \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}}\lambda_{\mathsf{m}} - \lambda_{\mathsf{n}} \left(1 - \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}} \right) \right).$$

The substitution of these expressions into (54) and (55) shows that $\alpha = \alpha(\lambda)$ and $\beta = \beta(\lambda)$ depend on λ, and that

$$\ln p_{\mathsf{m}} + 1 = -\alpha(\lambda) - \beta(\lambda)\lambda_{\mathsf{m}} \tag{56}$$

for $\mathsf{m} = \mathsf{n}$ and $\mathsf{m} = \mathsf{n}'$ according to (52) and (53). Now for every $\mathsf{j} \in \mathbb{N}^d$ with $\mathsf{j} \neq \mathsf{n}, \mathsf{n}'$ we have

$$\frac{\partial p_{\mathsf{n}}}{\partial p_{\mathsf{j}}} = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} (\lambda_{\mathsf{j}} - \lambda_{\mathsf{n}'}), \tag{57}$$

$$\frac{\partial p_{\mathsf{n}'}}{\partial p_{\mathsf{j}}} = (\lambda_{\mathsf{n}'} - \lambda_{\mathsf{n}})^{-1} (\lambda_{\mathsf{n}} - \lambda_{\mathsf{j}}). \tag{58}$$

Furthermore, let us define

$$\overline{\mathsf{S}} := p_{\mathsf{n}} \ln p_{\mathsf{n}} + p_{\mathsf{n}'} \ln p_{\mathsf{n}'} + \sum_{\mathsf{m}\in\mathbb{N}^d,\ \mathsf{m}\neq\mathsf{n},\ \mathsf{n}'} p_{\mathsf{m}} \ln p_{\mathsf{m}} \tag{59}$$

where p_{n} and $p_{\mathsf{n}'}$ are given by the above expressions. From (57) and (58) followed by the use of (54) and (55) we get

$$\begin{aligned}\frac{\partial \overline{\mathsf{S}}}{\partial p_{\mathsf{j}}} &= \left(\lambda_{\mathsf{n}'}-\lambda_{\mathsf{n}}\right)^{-1}\left(\left(\lambda_{\mathsf{j}}-\lambda_{\mathsf{n}'}\right)\left(\ln p_{\mathsf{n}}+1\right)+\left(\lambda_{\mathsf{n}}-\lambda_{\mathsf{j}}\right)\left(\ln p_{\mathsf{n}'}+1\right)\right)+\ln p_{\mathsf{j}}+1 \\ &= \alpha\left(\lambda\right)+\beta\left(\lambda\right)\lambda_{\mathsf{j}}+\ln p_{\mathsf{j}}+1, \end{aligned} \tag{60}$$

so that $\frac{\partial \overline{\mathsf{S}}}{\partial p_{\mathsf{j}}}=0$ if, and only if,

$$\ln p_{\mathsf{j}}+1=-\alpha\left(\lambda\right)-\beta\left(\lambda\right)\lambda_{\mathsf{j}}.$$

We now combine this with (56) to conclude that for every choice of probabilities which satisfy the hypotheses of the theorem and which annihilate (60) we have

$$\ln p_{\mathsf{m}}+1=-\alpha\left(\lambda\right)-\beta\left(\lambda\right)\lambda_{\mathsf{m}}$$

for every $\mathsf{m}\in\mathbb{N}^d$, that is,

$$p_{\mathsf{m}}=\exp\left[-\left(1+\alpha\left(\lambda\right)\right)\right]\exp\left[-\beta\left(\lambda\right)\lambda_{\mathsf{m}}\right].$$

Consequently, since $\lambda_{\mathsf{m}}\to+\infty$ as $|\mathsf{m}|\to+\infty$ we obtain (47) and (48) from (45) and (46), respectively, where $\beta\left(\lambda\right)>0$ and $\beta\in[\beta(\lambda),+\infty)$. This proves Statement (a) and gives (49) for every $\mathsf{m}\in\mathbb{N}^d$ along with (50).

In order to prove the remaining part of Statement (b) we now choose a large enough $\mathsf{M}:=(M_{1,\dots,}M_d)\in\mathbb{N}^d$ in such a way that $n_j<M_j$ and $n_j'<M_j$ for every $j\in\{1,\dots,d\}$, where n_j and n_j' are the components of n and n', respectively. We then consider the partial sums

$$\begin{aligned}\mathsf{S}_{\mathsf{M}} &:= \sum_{\mathsf{m}=(0,\dots,0)}^{\mathsf{M}} p_{\mathsf{m}}\ln\left(\frac{1}{p_{\mathsf{m}}}\right), \end{aligned} \tag{61}$$

$$\mathcal{Z}_{\mathsf{M}}(\beta) := \sum_{\mathsf{m}=(0,\dots,0)}^{\mathsf{M}} \exp\left[-\beta\lambda_{\mathsf{m}}\right]$$

of (43) and (47), respectively, where $m_j\in\{0,\dots,M_j\}$ for each j. In a similar way we define $\overline{\mathsf{S}}_{\mathsf{M}}$ from (59), so that the partial derivatives $\frac{\partial \overline{\mathsf{S}}_{\mathsf{M}}}{\partial p_{\mathsf{j}}}$ given by (60) now define a genuine finite-dimensional gradient. Therefore, the fact that (61) satisfies

$$\mathsf{S}_{\mathsf{M}}\leq\mathsf{S}_{\mathsf{M},\,\max} \tag{62}$$

where $\mathsf{S}_{\mathsf{M},\,\max}$ is evaluated by means of

$$p_{\mathsf{m},\,\mathsf{M}}:=\mathcal{Z}_{\mathsf{M}}^{-1}(\beta(\lambda))\exp\left[-\beta(\lambda)\lambda_{\mathsf{m}}\right]$$

follows from well-known considerations (see, e.g., Section 8 in Chapter 4 of [3]). Letting $|\mathsf{M}|\to+\infty$ in (62) then proves the desired result.

The differentiability of $\beta \mapsto \mathcal{Z}(\beta)$ at $\beta = \beta(\lambda)$ under the stated conditions as well as (51) are consequences of elementary arguments and of the direct substitution of (49) into (43). $\blacksquare$

REMARKS. (1) As is the case in Quantum Statistical Mechanics, Relation (46) has to be viewed as a further restriction on the class of admissible probabilities, which of course must be such that $p_{\mathsf{m}}\lambda_{\mathsf{m}} \to 0$ as $|\mathsf{m}| \to +\infty$. Furthermore, the choice of the preassigned value λ must be consistent with the nature of those eigenvalues. Thus if for instance $\lambda_{\mathsf{m}} \geq 0$ for every m, one must then impose $\lambda \in \mathbb{R}^+$ for (46) to make sense. We present an example of that kind in Section 3, where we also have $\sum_{\mathsf{m}\in\mathbb{N}^d} \exp\left[-\beta\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}} < +\infty$ for every $\beta \in (0, +\infty)$.

(2) Whereas the preceding considerations may be viewed as describing equilibrium situations in that (43) does not depend explicitly on time, there are many other entropy functionals associated with non-equilibrium situations which we may associate with Bernstein processes, for instance

$$\mathsf{S}_{\mathsf{m}}(t) := \int_{D\times D} d\mathsf{x}d\mathsf{y}w_{\mathsf{m}}(\mathsf{x}, t; \mathsf{y}, 0) \ln\left(\frac{1}{w_{\mathsf{m}}(\mathsf{x}, t; \mathsf{y}, 0)}\right) \tag{63}$$

in case of the Markovian processes of Theorem 1, where w_{m} given by (14) satisfies a specific Kolmogorov or Fokker-Planck equation. We defer the derivation of such equations, the analyses of the related entropy functionals such as (63) and their consequences to a separate publication.

We devote the next section to illustrating some of the above results.

3 A hierarchy of Bernstein processes in a two-dimensional disk

We consider here forward-backward problems of the form (2)-(3) with $V = 0$ identically, defined in the open two-dimensional disk of radius one centered at the origin, so that Hypothesis (H_1) trivially holds. We limit ourselves to an illustration of a few properties listed in the preceding section regarding Bernstein processes generated by certain radially symmetric solutions to such problems. Thus, we first switch to polar coordinates and start out with the hierarchy

$$\begin{aligned}
\partial_t u(r,t) &= \frac{1}{2}\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)u(r,t), \quad (r,t) \in (0,1] \times (0,T]\,, \\
u(r,0) &= \varphi_{0,\mathsf{m}}(r), \quad r \in [0,1]\,, \\
\partial_r u(1,t) &= 0, \quad t \in [0,T]
\end{aligned} \tag{64}$$

and

$$\begin{aligned}
-\partial_t v(r,t) &= \frac{1}{2}\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)v(r,t), \quad (r,t) \in (0,1] \times [0,T)\,, \\
v(r,0) &= \psi_{T,\mathsf{m}}(r), \quad r \in [0,1]\,, \\
\partial_r v(1,t) &= 0, \quad t \in [0,T]\,.
\end{aligned} \tag{65}$$

In this case the index $\mathsf{m} \in \mathbb{N}$ labels the discrete spectrum of the radial part of Neumann's Laplacian $-\frac{1}{2}\Delta_{\mathsf{x}}$ on the disk, which consists exclusively of eigenvalues $\lambda_{\mathsf{m}} \geq 0$ determined by the condition

$$J_1\left(\sqrt{2\lambda_{\mathsf{m}}}\right) = 0 \tag{66}$$

where J_1 stands for the Bessel function of the first kind of order one. For convenience we order these eigenvalues as

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < ..... , \tag{67}$$

and recall that there exists a finite constant $c > 0$ such that

$$c\,(\mathsf{m}-1)^2 < \lambda_{\mathsf{m}} < c\,(\mathsf{m}+1)^2 \tag{68}$$

for every $\mathsf{m} \in \mathbb{N}^+$. Moreover, the corresponding orthonormal basis $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}}$ of eigenfunctions in the space of all complex-valued, square-integrable functions with respect to the measure rdr on $(0,1)$ is given by

$$\mathsf{f}_{\mathsf{m}}(r) = \frac{\sqrt{2}}{\left|J_0\left(\sqrt{2\lambda_{\mathsf{m}}}\right)\right|} J_0\left(\sqrt{2\lambda_{\mathsf{m}}}r\right) \tag{69}$$

where J_0 stands for the Bessel function of the first kind of order zero. All these properties follow from standard Sturm-Liouville theory and from related properties of Bessel functions (it is worth recalling here that (66) is Neumann's boundary condition at $r = 1$ for the problem under consideration since $J_0' = -J_1$, see, e.g., Section 40 in Chapter VII of [20], and that the factor two in (66) and (69) is due to the factor one-half in (64)-(65)). For every $\mathsf{m} \in \mathbb{N}$ let us now choose the initial-final data as

$$\varphi_{0,\mathsf{m}}(r) = \begin{cases} \frac{1}{\pi} \;\text{ for } \mathsf{m} = 0, \\ \frac{1}{\pi}\left(1 + J_0\left(\sqrt{2\lambda_{\mathsf{m}}}r\right)\right) \;\text{ for } \mathsf{m} \in \mathbb{N}^+ \end{cases} \tag{70}$$

and

$$\psi_{T,\mathsf{m}}(r) = 1, \tag{71}$$

respectively. It follows from (70) and (71) that Hypothesis (H_2) holds, a consequence of elementary properties of J_0 including its uniform boundedness in case of (70). The corresponding solutions to (64) and (65) then read

$$u_{\mathsf{m}}(r,t) = \begin{cases} \frac{1}{\pi} \;\text{ for } \mathsf{m} = 0, \\ \frac{1}{\pi}\left(1 + \exp\left[-t\lambda_{\mathsf{m}}\right] J_0\left(\sqrt{2\lambda_{\mathsf{m}}}r\right)\right) \;\text{ for } \mathsf{m} \in \mathbb{N}^+ \end{cases} \tag{72}$$

and

$$v_{\mathsf{m}}(r,t) = 1 \tag{73}$$

for each $r \in [0,1]$ and every $t \in [0,T]$, respectively. Moreover, as a consequence of (9), (70) and (73) we also have

$$\int_{|\mathsf{x}|<1} \mathsf{dx}\varphi_{\mathsf{m},0}(\mathsf{x}) \int_{|\mathsf{x}|<1} \mathsf{dy} g(\mathsf{x},T,\mathsf{y})\psi_{\mathsf{m},T}(\mathsf{y})$$

$$= \int_{|\mathsf{x}|<1} \mathsf{dx}\varphi_{\mathsf{m},0}(\mathsf{x}) = 2\pi \int_0^1 dr r \varphi_{\mathsf{m},0}(r) = 1$$

so that (7) is verified. We may therefore apply all the results of the preceding section to the present situation, some of which we state in the following proposition where

$$\mathbb{D} := \left\{ \mathsf{x} \in \mathbb{R}^2 : |\mathsf{x}| < 1 \right\}.$$

Proposition 2. *For each $\mathsf{m} \in \mathbb{N}$ let μ_{m} be the measure of the form (6) with the initial-final data given by (70) and (71), respectively. Then, there exists a $\overline{\mathbb{D}}$-valued, non-stationary Markovian Bernstein process $Z^{\mathsf{m}}_{\tau\in[0,T]}$ such that the following properties hold:*

(a) For each Borel subset $F \subseteq \overline{\mathbb{D}}$ of Lebesgue measure $|F|$ and for every $t \in [0,T]$ we have

$$\mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_t \in F\right) = \begin{cases} \frac{|F|}{\pi} \quad \text{for } \mathsf{m} = 0, \\ \frac{1}{\pi}\left(|F| + \exp\left[-t\lambda_{\mathsf{m}}\right] \int_F \mathsf{dx} J_0\left(\sqrt{2\lambda_{\mathsf{m}}}\,|\mathsf{x}|\right)\right) \quad \text{for } \mathsf{m} \in \mathbb{N}^+. \end{cases}$$

Thus, the function $t \to \mathbb{P}_{\mu_{\mathsf{m}}}\left(Z^{\mathsf{m}}_t \in F\right)$ is non-increasing on $[0,T]$.

(b) For each bounded Borel measurable function $b : \overline{\mathbb{D}} \mapsto \mathbb{C}$ and every $t \in [0,T]$ we have

$$\mathbb{E}_{\mu_{\mathsf{m}}}\left(b(Z^{\mathsf{m}}_t)\right) = \begin{cases} \frac{1}{\pi}\int_{|\mathsf{x}|<1} \mathsf{dx} b(\mathsf{x}) \quad \text{for } \mathsf{m} = 0, \\ \frac{1}{\pi}\int_{|\mathsf{x}|<1} \mathsf{dx} b(\mathsf{x})\left(1 + \exp\left[-t\lambda_{\mathsf{m}}\right] J_0\left(\sqrt{2\lambda_{\mathsf{m}}}\,|\mathsf{x}|\right)\right) \quad \text{for } \mathsf{m} \in \mathbb{N}^+. \end{cases}$$

(c) Let $\lambda \in \mathbb{R}^+$ be given. Then, the process $\bar{Z}_{\tau\in[0,T]}$ of maximal entropy within $\overline{\mathbb{D}}$ in the sense of Theorem 4 is obtained by averaging the $Z^{\mathsf{m}}_{\tau\in[0,T]}$ with probabilities of the form

$$p_{\mathsf{m}} = \mathcal{Z}^{-1}(\beta(\lambda)) \exp\left[-\beta(\lambda)\lambda_{\mathsf{m}}\right]$$

where $\mathcal{Z}(\beta(\lambda))$ is given by (47). Moreover, $\bar{Z}_{\tau\in[0,T]}$ is Markovian and its entropy may be evaluated from (51).

The proof is a direct application of the corresponding formulae in Section 2 combined with those of this section. We note that we must have $\lambda > 0$ for the preassigned value in Statement (c) since $\lambda_{\mathsf{m}} \geq 0$ for every $\mathsf{m} \in \mathbb{N}$ according to (67). We also have $\sum_{\mathsf{m}\in\mathbb{N}^d}\exp\left[-\beta\lambda_{\mathsf{m}}\right]\lambda_{\mathsf{m}}<+\infty$ for every $\beta\in(0,+\infty)$ as a consequence of (68), so that expression (51) may indeed be applied in this case.

Remarks. (1) Although the mixed processes obtained by the averaging method described in Section 2 are not Markovian in general (see the remark following the proof of Theorem 2), the reason why $\bar{Z}_{\tau\in[0,T]}$ possesses the Markov property is due to our choice of the final condition (71), which implies that the averaged joint density (21) becomes

$$\overline{\mu}\left(\mathsf{x},\mathsf{y}\right)=g(\mathsf{x},T,\mathsf{y})\sum_{\mathsf{m}\in\mathbb{N}}p_{\mathsf{m}}\varphi_{\mathsf{m},0}\left(\mathsf{x}\right)$$

where $\varphi_{\mathsf{m},0}$ is given by (70). Indeed, the preceding relation is then of the form (17) with an obvious choice for ν_0 and ν_1. For more details and examples regarding the time evolution of Bernstein processes that possess the Markov property we refer the reader to [17].

(2) We can obtain similar results for radially symmetric forward-backward problems of the form (2)-(3) with $V=0$ defined in the open ball $\mathsf{B}^d\subset\mathbb{R}^d$ of radius one centered at the origin where $d\geq 3$. The eigenfunctions of the radial part of the Laplacian then involve the Bessel function $J_{\frac{d}{2}-1}$ rather than J_0, while Neumann's boundary condition is expressed in terms of $J_{\frac{d}{2}}$ instead of J_1 (the case $d=1$ can be dealt with directly in terms of trigonometric functions). However, the corresponding formulae for the probabilities and the expectation values of the underlying processes become much more involved.

Appendix A. A class of time-dependent statistical operators

In this appendix we define and investigate statistical operators which are more general than that defined by (33), in view of getting expressions such as (42) for the averaged processes of Theorem 2 which are as a rule non-stationary. This, however, requires some additional structure. Let us write

$$\exp\left[-t\mathcal{H}\right]f\left(.\right):=\begin{cases} f(.) \quad \text{if } t=0, \\ \int_D \mathsf{d}\mathsf{y}g(.,t,\mathsf{y})f(\mathsf{y}) \quad \text{if } t\in(0,T] \end{cases} \tag{74}$$

for the positivity preserving, symmetric semigroup generated by (1) on $L^2(D)$, where g is given by (4). In addition to Hypothesis (H_2) regarding the initial-final data we shall now impose the following requirement:

(H_3) The sequences $\left(\varphi_{\mathsf{m},0}\right)_{\mathsf{m}\in\mathbb{N}^d}$ and $\left(\exp\left[-T\mathcal{H}\right]\psi_{\mathsf{m},T}\right)_{\mathsf{m}\in\mathbb{N}^d}$ form a biorthonormal system in $L^2(D)$, that is,

$$\left(\varphi_{\mathsf{m},0},\exp\left[-T\mathcal{H}\right]\psi_{\mathsf{n},T}\right)_2=\delta_{\mathsf{m},\mathsf{n}}$$

for all $\mathsf{m},\mathsf{n}\in\mathbb{N}^d$.

First we have:

Lemma 2. *Let us assume that Hypotheses* (H_2) *and* (H_3) *hold. Then the sequences of solutions* $(u_{\mathsf{m}}(.,t))_{\mathsf{m}\in\mathbb{N}^d}$, $(v_{\mathsf{m}}(.,t))_{\mathsf{m}\in\mathbb{N}^d}$ *given by (8) and (9), respectively, form a biorthonormal system in* $L^2(D)$ *for every* $t \in [0,T]$.

Proof. From (8), (9) and (74) we have

$$\begin{aligned} & (u_{\mathsf{m}}(.,t), v_{\mathsf{n}}(.,t))_2 \\ = \; & \left(\exp\left[-t\mathcal{H}\right]\varphi_{\mathsf{m},0}, \exp\left[-(T-t)\mathcal{H}\right]\psi_{\mathsf{n},T}\right)_2 \\ = \; & \left(\varphi_{\mathsf{m},0}, \exp\left[-T\mathcal{H}\right]\psi_{\mathsf{n},T}\right)_2 = \delta_{\mathsf{m},\mathsf{n}} \end{aligned} \tag{75}$$

from the symmetry of the semigroup and (H_3), for all $\mathsf{m},\mathsf{n} \in\mathbb{N}^d$. ■

We then define

$$\mathcal{R}\left(t\right)f := \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\left(f, u_{\mathsf{m}}(.,t)\right)_2 v_{\mathsf{m}}(.,t) \tag{76}$$

for each $f \in L^2(D)$ and every $t \in [0,T]$. The following result shows that (76) possesses several properties similar to those stated in Proposition 1:

Proposition A.1. *Let us assume that Hypotheses* (H_1), (H_2) *and* (H_3) *hold. Then the following statements are valid:*

(a) Expression (76) defines a linear trace-class operator in $L^2\left(D\right)$ *such that*

$$\operatorname{Tr}\mathcal{R} = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} = 1 \tag{77}$$

and

$$\operatorname{Tr}\mathcal{R}^2 = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}^2 \leq 1 \tag{78}$$

for every $t \in [0,T]$.

(b) The eigenvalue equation

$$\mathcal{R}\left(t\right)v_{\mathsf{m}}(.,t) = p_{\mathsf{m}} v_{\mathsf{m}}(.,t) \tag{79}$$

holds for each $\mathsf{m} \in\mathbb{N}^d$ *and every* $t \in [0,T]$.

(c) If B *is a linear bounded self-adjoint operator on* $L^2(D)$ *we have*

$$\operatorname{Tr}\left(\mathcal{R}(t)B\right) = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\left(Bu_{\mathsf{m}}(.,t), v_{\mathsf{m}}(.,t)\right)_2 \tag{80}$$

for every $t \in [0,T]$. *In particular, if* B *is the same multiplication operator as in Part (c) of Proposition 1 we have*

$$\operatorname{Tr}\left(\mathcal{R}(t)B\right) = \mathbb{E}_{\bar{\mu}}\left(b(\bar{Z}_t)\right) \tag{81}$$

where the right-hand side of (81) is given by (23).

Proof. The proof of the trace-class property is similar to that given in Proposition 1. Thus, we first remark that there exists a finite constant $c > 0$ such that the estimates

$$\|u_{\mathsf{m}}(.,t)\|_2 \leq c$$

and

$$\|v_{\mathsf{m}}(.,t)\|_2 \leq c$$

hold uniformly in m and t as a consequence of (4), (8), (9) and Hypothesis (H_2), which makes a linear bounded operator out of (76) on $L^2(D)$. The relevant auxiliary function for the remaining part of the argument is then given by

$$a(\mathsf{m},\mathsf{n},t) := p_{\mathsf{m}}\left(\mathsf{h}_{\mathsf{n}}, u_{\mathsf{m}}(.,t)\right)_2 \left(v_{\mathsf{m}}(.,t), \mathsf{h}_{\mathsf{n}}\right)_2 \tag{82}$$

for every $t \in [0,T]$, with $(\mathsf{h}_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}^d}$ an arbitrary orthogonal basis as before. It is indeed easily seen that the properties of (82) are similar to those of (39), the key point in getting (77) and (78) being the biorthogonality relation (75) which replaces the orthogonality properties of $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$. The same observation applies for the proof of (79), while (80) follows from the relation

$$(\mathcal{R}(t)B\mathsf{h}_{\mathsf{n}}, \mathsf{h}_{\mathsf{n}})_2 = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\left(\mathsf{h}_{\mathsf{n}}, Bu_{\mathsf{m}}(.,t)\right)_2 \left(v_{\mathsf{m}}(.,t), \mathsf{h}_{\mathsf{n}}\right)_2$$

valid for every $\mathsf{n} \in \mathbb{N}^d$, Relation (80) then implying (81). ■

Remarks. (1) As a consequence of (74) and elementary spectral theory, it is plain that the forward and backward solutions (8) and (9) become

$$u_{\mathsf{m}}(.,t) = \exp\left[-tE_{\mathsf{m}}\right]\mathsf{f}_{\mathsf{m}}$$

and

$$v_{\mathsf{m}}(.,t) = \exp\left[tE_{\mathsf{m}}\right]\mathsf{f}_{\mathsf{m}},$$

respectively, for initial-final data given by (24). The substitution of these expressions into (76) then gives (33), so that the former operators are indeed time-dependent generalizations of the latter. Moreover, thanks to Statement (a) of Proposition A.1 the classification of processes as pure or mixed according to Definition 2 still holds for the more general form (76).

(2) We may write (76) as

$$\mathcal{R}(t) f = \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}}\mathcal{P}_{\mathsf{m}}(t) f$$

where the operators given by

$$\mathcal{P}_{\mathsf{m}}(t) f := \left(f, u_{\mathsf{m}}(.,t)\right)_2 v_{\mathsf{m}}(.,t)$$

satisfy

$$\mathcal{P}_{\mathsf{m}}^2(t) = \mathcal{P}_{\mathsf{m}}(t)$$

for each $\mathsf{m} \in \mathbb{N}^d$ and every $t \in [0,T]$ as a consequence of the biorthogonality property. Therefore, we may view $\mathcal{R}(t)$ as a statistical mixture of oblique projections as the $\mathcal{P}_{\mathsf{m}}(t)$ are not self-adjoint in the general case. In fact, the adjoint of (76) in $L^2(D)$ is obtained by swapping the rôles of $u_{\mathsf{m}}(.,t)$ and $v_{\mathsf{m}}(.,t)$, that is,

$$\mathcal{R}^*(t)\, f := \sum_{\mathsf{m}\in\mathbb{N}^d} p_{\mathsf{m}} \left(f, v_{\mathsf{m}}(.,t)\right)_2 u_{\mathsf{m}}(.,t). \tag{83}$$

We note that (83) enjoys the very same properties as (76), with the exception of (79) which has to be replaced by

$$\mathcal{R}^*(t)\, u_{\mathsf{m}}(.,t) = p_{\mathsf{m}} u_{\mathsf{m}}(.,t).$$

We could therefore have chosen (83) as statistical operators instead of (76). Finally, we remark that $\mathcal{R}(t)$ and $\mathcal{R}^*(t)$ both involve (8) and (9), in agreement with the fact that there are two time directions in the theory from the outset.

(3) It is reasonable to ask whether it is always possible to choose initial-final data so that Hypothesis (H_3) holds. The answer is affirmative provided the functions $\exp[-T\mathcal{H}]\,\psi_{\mathsf{m},T}$ remain close to the orthonormal basis $(\mathsf{f}_{\mathsf{m}})_{\mathsf{m}\in\mathbb{N}^d}$ in some very specific $L^2(D)$-sense. This follows from a direct application of the generalization of a theorem by Paley and Wiener as stated in Section 86 of Chapter V in [12]. In that case the sequences $\left(\varphi_{\mathsf{m},0}\right)_{\mathsf{m}\in\mathbb{N}^d}$ and $\left(\exp[-T\mathcal{H}]\,\psi_{\mathsf{m},T}\right)_{\mathsf{m}\in\mathbb{N}^d}$ form a *complete* biorthogonal system.

Appendix B. On the existence of Bernstein processes and their relation with Schrödinger's problem and Optimal Transport Theory

The typical construction of a Bernstein process with state space $\overline{D}$ requires a probability measure μ on $\mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$ and a transition function Q, as is the case for Markov processes. We provide below a general theorem which is a direct consequence of a more abstract construction carried out in [8], or with a more analytical flavor in [18], to which we refer the reader for details. The theorem shows that all the basic quantities that characterize a Bernstein process can be expressed exclusively in terms of μ and the heat kernel g, which is all we needed in the preceding sections.

Since there are two time directions provided by (2)-(3), the natural choice for the transition function we alluded to is

$$Q\left(\mathsf{x}, t; F, r; \mathsf{y}, s\right) := \int_F \mathsf{dz}\, q\left(\mathsf{x}, t; \mathsf{z}, r; \mathsf{y}, s\right) \tag{84}$$

for every $F \in \mathcal{B}(\overline{D})$, where

$$q\left(\mathsf{x}, t; \mathsf{z}, r; \mathsf{y}, s\right) := \frac{g(\mathsf{x}, t-r, \mathsf{z})g(\mathsf{z}, r-s, \mathsf{y})}{g(\mathsf{x}, t-s, \mathsf{y})}. \tag{85}$$

Both functions are well defined and positive for all $\mathsf{x}, \mathsf{y}, \mathsf{z} \in \mathbb{R}^d$ and all r, s, t satisfying $r \in (s,t) \subset [0,T]$ by virtue of (4), and moreover the normalization condition

$$Q\left(\mathsf{x}, t; D, r; \mathsf{y}, s\right) = 1$$

holds as a consequence of the semigroup composition law for g. The result we have in mind is the following:

Theorem B.1. *Let μ be a probability measure on $\mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$, and let Q be given by (84). Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ supporting an $\overline{D}$-valued Bernstein process $Z_{\tau\in[0,T]}$ such that the following properties are valid:*

(a) The function Q is the two-sided transition function of $Z_{\tau\in[0,T]}$ in the sense that

$$\mathbb{P}_\mu \left(Z_r \in F \,|Z_s, Z_t \right) = Q \left(Z_t, t; F, r; Z_s, s \right) \tag{86}$$

for each $F \in \mathcal{B}(\overline{D})$ and all r, s, t satisfying $r \in (s,t) \subset [0,T]$. Moreover,

$$\mathbb{P}_\mu \left(Z_0 \in F_0, Z_T \in F_T \right) = \mu \left(F_0 \times F_T \right) \tag{87}$$

for all $F_0, F_T \in \mathcal{B}(\overline{D})$, that is, μ is the joint probability distribution of Z_0 and Z_T. In particular, the marginal distributions are given by

$$\mathbb{P}_\mu \left(Z_0 \in F \right) = \mu \left(F \times \overline{D} \right) \tag{88}$$

and

$$\mathbb{P}_\mu \left(Z_T \in F \right) = \mu \left(\overline{D} \times F \right) \tag{89}$$

for each $F \in \mathcal{B}(\overline{D})$, respectively.

(b) For every $n \in \mathbb{N}^+$ the finite-dimensional distributions of the process are given by

$$\begin{aligned} & \mathbb{P}_\mu \left(Z_{t_1} \in F_1, ..., Z_{t_n} \in F_n \right) \\ = \; & \int_{D\times D} \frac{\mathsf{d}\mu(\mathsf{x},\mathsf{y})}{g(\mathsf{x},T,\mathsf{y})} \int_{F_1} \mathsf{dx}_1 ... \int_{F_n} \mathsf{dx}_n \\ & \times \prod_{k=1}^{n} g \left(\mathsf{x}_k, t_k - t_{k-1}, \mathsf{x}_{k-1} \right) \times g \left(\mathsf{y}, T - t_n, \mathsf{x}_n \right) \end{aligned} \tag{90}$$

for all $F_1, ..., F_n \in \mathcal{B}(\overline{D})$ and all $t_0 = 0 < t_1 < ... < t_n < T$, where $\mathsf{x}_0 = \mathsf{x}$. In particular we have

$$\begin{aligned} & \mathbb{P}_\mu \left(Z_t \in F \right) \\ = \; & \int_{D\times D} \frac{\mathsf{d}\mu(\mathsf{x},\mathsf{y})}{g(\mathsf{x},T,\mathsf{y})} \int_{F} \mathsf{dz} g \left(\mathsf{x}, t, \mathsf{z} \right) g \left(\mathsf{z}, T - t, \mathsf{y} \right) \end{aligned} \tag{91}$$

for each $F \in \mathcal{B}(\overline{D})$ and every $t \in (0,T)$.

(c) $\mathbb{P}_\mu$ is the only probability measure leading to the above properties.

As we saw in Section 2, Theorems 1, 2 and 3 were obtained by substituting the respective measures in the formulae of Theorem B.1.

We conclude this appendix with a very brief remark which establishes the connection between Markovian Bernstein processes on the one hand, Schrödinger's problem and Optimal Transport Theory on the other hand. From Section 2 we know that in the Markovian case the relevant probability measures to be substituted into the formulae of Theorem B.1 are necessarily of the form

$$\mu(G) = \int_G \mathsf{dxdy}\varphi_0(\mathsf{x})g(\mathsf{x},T,\mathsf{y})\psi_T(\mathsf{y})$$

where $G \in \mathcal{B}(\overline{D}) \times \mathcal{B}(\overline{D})$, and where $\varphi_0 > 0$ and $\psi_T > 0$. In particular, the marginal distributions are

$$\mu(F \times \overline{D}) = \int_F \mathsf{dx}\varphi_0(\mathsf{x}) \int_D \mathsf{dy} g(\mathsf{x},T,\mathsf{y})\psi_T(\mathsf{y})$$

and

$$\mu(\overline{D} \times F) = \int_D \mathsf{dx}\varphi_0(\mathsf{x})g(\mathsf{x},T,\mathsf{y}) \int_F \mathsf{dy}\psi_T(\mathsf{y})$$

where $F \in \mathcal{B}(\overline{D})$, which gives rise to the respective densities

$$\mu_0(\mathsf{x}) := \varphi_0(\mathsf{x}) \int_D \mathsf{dy} g(\mathsf{x},T,\mathsf{y})\psi_T(\mathsf{y}) \tag{92}$$

and

$$\mu_T(\mathsf{y}) := \psi_T(\mathsf{y}) \int_D \mathsf{dx}\varphi_0(\mathsf{x})g(\mathsf{x},T,\mathsf{y}). \tag{93}$$

Thus, the considerations of this article show that these marginal densities are entirely determined by the initial-final data once the heat kernel is known. It is, however, the inverse point of view that prevails in Schrödinger's problem and in the related Optimal Transport Theory, which *first* amounts to prescribing μ_0 and μ_T and *then* consider (92) and (93) as a nonlinear inhomogeneous system of integral equations in the two unknowns φ_0 and ψ_T. That is actually what was developed by Schrödinger in the last part of [13] by using entropy arguments. Moreover, given μ_0 and μ_T continuous, a very general existence and uniqueness result for the pair (φ_0, ψ_T) satisfying (92) and (93) was proved in [2]. In the context of Optimal Transport Theory, arguments that allow the minimization of the so-called cost functions using entropy related methods are often used. We refer the reader fro instance to [9] and to the references therein for details.

Acknowledgements. The author would like to thank the Fundação para a Ciência e Tecnologia (FCT) of the Portuguese Government for its financial support under Grant PTDC/MAT-STA/0975/2014. He is also indebted to L. Monsaingeon and J.-C. Zambrini for stimulating discussions regarding some aspects of this article. Parts of this work were presented as an invited contributed talk at the 2nd Conference of the European Physical Society Statistical and Nonlinear Physics Division, which took place at Nordita, Stockholm, in May of 2019. The author would like to thank the organizers for their very kind invitation.

References

[1] Bernstein, S., *Sur les liaisons entre les grandeurs aléatoires,* in: Verhandlungen des Internationalen Mathematikerkongress **1** (1932) 288-309.

[2] Beurling, A., *An automorphism of product measures,* Annals of Mathematics **72**, 189-200 (1960).

[3] Brillouin, L., *Science and Information Theory,* Academic Press, New York (1967).

[4] Cruzeiro, A. B., Wu, L., Zambrini, J. C., *Bernstein processes associated with Markov processes,* in: Stochastic Analysis and Mathematical Physics (R. Rebolledo, editor), Birkhäuser, Basel (2000).

[5] Edmunds, D. E., Evans, W. D., *Spectral Theory and Differential Operators*, Oxford Mathematical Monographs, Clarendon Press, Oxford (1987).

[6] Gohberg, I. C., Krein, M. G., *Introduction to the Theory of Linear Nonselfadjoint Operators in Hilbert Space,* Translations of Mathematical Monographs **18**, American Mathematical Society, Providence, (1969).

[7] Itô, K., *Stochastic Processes,* in: Lectures given at Aarhus University (O. E. Barndorff-Nielsen and K. Sato, editors), Springer, New York (2004).

[8] Jamison, B., *Reciprocal processes,* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **30** (1974) 65-86.

[9] Léonard, C., *A survey of the Schrödinger problem and some of its connections with optimal transport,* Discrete and Continuous Dynamical Systems, Series A **34** (2014) 1533-1574.

[10] Léonard, C., Roelly, S., Zambrini, J. C., *Reciprocal processes. A measure-theoretical point of view,* Probability Surveys **11** (2014) 237-269.

[11] Loève, M., *Probability Theory,* Van Nostrand Publishing Company, Princeton (1963).

[12] Riesz, F., Nagy, B., Sz., *Functional Analysis,* Dover Books on Mathematics, Dover (1990).

[13] Schrödinger, E., *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique,* Annales de l'Institut Henri Poincaré **2** (1932) 269-310.

[14] Vainberg, M, M., *Variational Methods for the Study of Nonlinear Operators,* Holden-Day Series in Mathematical Physics, Holden-Day, Inc., San Francisco (1964).

[15] Villani, C., *Optimal Transport: Old and New,* Grundlehren der Mathematischen Wissenschaften **338**, Springer, New York (2009).

[16] von Neumann, J., *Mathematical Foundations of Quantum Mechanics,* Princeton Landmarks in Mathematics Series, Princeton University Press, Princeton (1996).

[17] Vuillermot, P.-A., *On the time evolution of Bernstein processes associated with a class of parabolic equations,* Discrete and Continuous Dynamical Systems Series B, **23** (2018) 1073-1090.

[18] Vuillermot, P.-A., Zambrini, J. C., *Bernstein diffusions for a class of linear parabolic partial differential equations,* Journal of Theoretical Probability **27** (2014) 449-492.

[19] Vuillermot, P.-A., Zambrini, J. C., *On Bernstein processes generated by hierarchies of linear parabolic systems in* $\mathbb{R}^d$*,* arXiv: 1802.07077, (2018).

[20] Weinberger, H. F., *A First Course in Partial Differential Equations with Complex Variables and Transform Methods,* Dover Publications, Inc. New York (1965).

[21] Zambrini, J. C., *The research program of Stochastic Deformation (with a view toward Geometric Mechanics),* in: Stochastic Analysis, a Series of Lectures, Birkhäuser Progress in Probability book series **68**, Eds. R. Dalang, M. Dozzi, F. Flandoli, F. Russo (2015), pp. 359-393.