From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform

Jan Kloppenborg Møller[1,2], and Henrik Madsen[1]

December 14, 2010

Abstract

This report describes methods to eliminate state dependent diffusion terms in Stochastic Differential Equations (SDEs). Transformations that leave the diffusion term of SDEs constant is important for simulation, and estimation. It is important for simulation because the Euler approximation convergence rate is faster, and for estimation because the Extended Kalman Filter equations are easier to implement than higher order filters needed in the case of state dependent diffusion terms. The general class of transformations which leaves the diffusion term independent of the state is called the Lamperti transform. This note gives an example driven introduction to the Lamperti transform. The general applicability of the Lamperti transform is limited to univariate diffusion processes, but for a restricted class of multivariate diffusion processes Lamperti type transformations are available and the Lamperti transformation is discussed for both univariate and multivariate diffusion processes. Further some special attention is needed for time-inhomogeneous diffusion processes and these are discussed separately.

Key Words: Stochastic Differential Equations, Level dependent diffusion, Lamperti transform, Extended Kalman Filter.

[1]Corresponding author, E-mail: jkm@imm.dtu.dk, Tlf: +45 4525 3349

[2]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.

1 Introduction

Stochastic differential equations (SDE's) are attracting increasing attention, because physical processes in real life systems experience random forcing, due to model approximations and stochastic inputs, that cannot be captured by ordinary differential equations (ODE's). Such random forcing or internal noise can be captured by adding random noise in the ODE, and this leads to SDE formulations.

The formulation of SDE's is done by physical reasoning. This physical reasoning includes autocorrelation structures and physical constraints (such as mass balance considerations) captured by the diffusion term. The formulation and reasoning often results in structures where the noise (diffusion) term depends on one or more state variables. Structures where the diffusion term depend on the state of the system are difficult to handle in estimation procedure like the one implemented in CTSM[1] (Kristensen & Madsen , 2003; Kristensen et. al., 2004), since the Extended Kalman Filter (EKF) requires higher (than 1) order terms in order to make the filter approximations sufficiently accurate. Therefore transformations that can move (or remove) the state dependence from the diffusion term to the drift term are needed. Other estimation procedures (Iacus, 2008) also rely on the existence of transformations of this sort. Transformations to unit diffusion is often referred to as Lamperti transform.

Further it is often recommended (Iacus, 2008) to use the Lamperti transformation before simulations. State dependent diffusion can together with structures in the drift term impose restrictions on the state space, e.g. processes that exist on the positive real axis only, like the Black and Scholes model (geometric Brownian motion). Estimation of such systems is not numerically stable if combined with a observation equation that use these constraints (like the $\log$-transform), since estimation of the process may be zero (the geometric Brownian motion is strictly positive). However, after an appropriate transformation this process lives on the entire real axis and numerical problems on the boundary of the domain is avoided.

The results presented here seems to be well-known in more theoretical literature on SDE's (e.g. Luschgy, 2006), it is however hard to find papers, that explicitly deals with the construction of these kind of transformation in more applied settings. An exception is Nielsen & Madsen (2001), but comparing the results presented in that reference and the results presented here shows that the results in Nielsen & Madsen (2001) need corrections. Aït-Sahalia (2008) present transformations for a more general class of SDEs (referred to as reducible), these transformations are however more complicated to apply and we lose the generic formulations obtained in this report.

[1] `www.imm.dtu.dk/ctsm`

The report starts with a presentation of the general setting in Section 2. Results on one dimensional diffusion are given in Section 3, which is further divided into time independent (Section 3.1) and time dependent diffusion (Section 3.2). The theoretical properties do not differ much between the two cases, but for practical applications some notes are needed for the time dependent diffusion. The multivariate case is presented in Section 4. This part does not consider a split into time independent and time dependent diffusion, since the remarks on the one dimensional time dependent case applies equally to the multidimensional case. Finally Section 5 gives a short summary and discussion of the result presented.

2 The general setting

Itô processes (SDE's) which are partly observed in discrete time are referred to as the continuous-discrete time stochastic state space models (Jazwinski, 1970), and a general formulation is

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{1}$$

$$\boldsymbol{Y}_k = \boldsymbol{g}(\boldsymbol{X}_{t_k}, t_k, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, \boldsymbol{e}_k), \tag{2}$$

where $t \in \mathbb{R}_0$ is time, $\boldsymbol{w}_t \in \mathbb{R}^m$ is the standard Brownian motion, $\boldsymbol{X}_t \in \mathbb{R}^n$ is the state variable, $\boldsymbol{u}_t \in \mathbb{R}^q$ is the input, $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector, $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ is a vector function and $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times m}$ is the diffusion matrix. In the observation equation (2) $\boldsymbol{y} \in \mathbb{R}^l$ is the observations of state variable, $\boldsymbol{g} \in \mathbb{R}^l$ is the observation function and $\boldsymbol{e}_k \in \mathbb{R}^r$ is the observation error. The estimation problem is: Find $\hat{\boldsymbol{\theta}}$ such that

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}(S(\boldsymbol{\theta}, \mathcal{Y}_N)), \tag{3}$$

where S is some objective function and $\mathcal{Y}_N = \{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_N\}$ is the set of all observations. The obvious choice for S is to maximise the one-step transitions probabilities, i.e. the product of the probability density functions (pdf's) $p(Y_k|\mathcal{Y}_{k-1})$. This product is called the likelihood function (in practice we optimise the $\log$-likelihood). The likelihood can in principle be found by solving the Fokker-Planck equation (Gard, 1988; Klebaner, 2005) and using Bayes rule for updating. It is however unrealistic to solve the Fokker-Planck equation if the system equation (1) does not have a very simple form. The general situation is sketched in Figure 1, to obtain the transition probability we need to integrate the SDE Eq. (1) between observations, when an observation is available the information provided by this observation is used to form the reconstruction of the state, and the transition probability to the next observation is again obtained by integration.

One way to move forward is by approximating the transition probabilities by Gaussian pdf's, and transforming the observation equation such that the observation noise is (approximately) additive Gaussian. In order to calculate the likelihood function Extended Kalman Filters (EKFs) are often used, where the filter equations take complicated forms (higher order moments is needed and numerical solutions tend to be unstable if the diffusion term is a function of the state). It is therefore advisable to use transformations of the system equation (1) such that the diffusion is independent of the state. The transformation (ψ in (Figure 1) should form an equivalent relation between the input $\boldsymbol{u}_t$ and the output $\boldsymbol{Y}_k$ and the transformed system equation should depend on the same parameters ($\boldsymbol{t}heta$). Even if the main problem is estimation, the application is more general since it is well-known that simulations has better convergence rates (Iacus, 2008) if the system equation is independent of the states. The subject of this note is transformations of the system equations that leaves the diffusion of the transformed system equations independent of the state.

In the following we will restrict the analysis to $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$. There are two remarks about this 1) most derivations (except transformation to unity) generalise easily to the general case, and 2) in a weak solution sense (equality in distribution) this is not a restriction, since $\boldsymbol{\sigma}(\cdot)$ is only unique up to the (definite) "square root" of $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$. A small example can illustrate the last point.

Example 1 *Consider the SDE*

$$dX_t = adt + \sigma_1 dw_{1,t} + \sigma_2 dw_{2,t}; \quad X_0 = 0, \tag{4}$$

where a, σ_1 and σ_2 are real constants. The solution to (4) is

$$X_t = at + \sigma_1 w_{1,t} + \sigma_2 w_{2,t}, \tag{5}$$

which is a Gaussian distributed random variable with mean and variance equal to at and $\sigma_1^2 + \sigma_2^2$, respectively, but this is also the (weak) solution to

$$dX_t = adt + \sqrt{\sigma_1^2 + \sigma_2^2} dw_t; \quad X_0 = 0, \tag{6}$$

which illustrates that the uniqueness of the weak solution is only unique up to the square root of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. □

The implication will be discussed further for multivariate processes in Section 4. The term "weak solution" refer to equality in distribution, and strong solutions refer to path-wise equality (see Øksendal (2003) for further discussions on weak

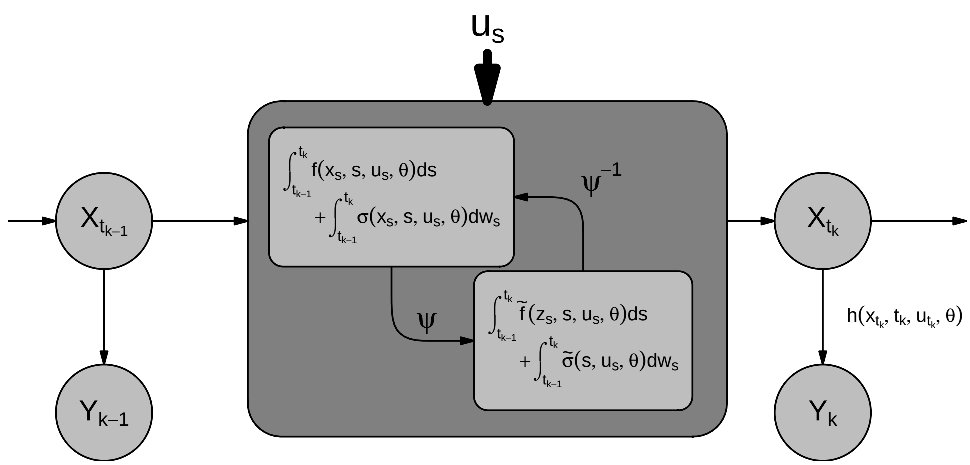

Figure 1: Conceptual diagram of the estimation problem, when an observation Y_k is available the state estimate of $X_{t_{k-1}}$ is updated by the provided information and used for integration of the state to form the prediction of the state X_{t_k}. There is an infinite number of equivalent relations between the input u_t and the output Y_k, the equivalence relation ψ gives a description with the same parameter, but $\tilde{\sigma}$ is independent of z_t.

and strong solutions). Clearly a strong solution is a weak solution, but a weak solution is not necessarily a strong solution (just consider Example 1). In this note we will refer to weak solutions (which might also be strong) as solutions. In likelihood estimation the only interest is weak solutions, since we optimise the distribution. In simulation studies tha main interest will often also be weak solution.

2.1 Notation and problem setting

This note is only concerned with the system equation and with the comments above the class of differential equations is restricted to

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{7}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$, $\boldsymbol{w}_t \in \mathbb{R}^n$ is the standard Brownian motion, and all other variable and functions are as explained below Eq. (1). This note deals with the problem; Find transformations $\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t)$ or $\tilde{\boldsymbol{Z}}_t = \tilde{\boldsymbol{\psi}}(\boldsymbol{X}_t, t)$ such that

$$d\boldsymbol{Z}_t = \widetilde{\boldsymbol{f}}(\boldsymbol{Z}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \widetilde{\boldsymbol{\sigma}}(t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{8}$$

or

$$d\tilde{\boldsymbol{Z}}_t = \widetilde{\boldsymbol{f}}_{\tilde{Z}}(\tilde{\boldsymbol{Z}}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + d\boldsymbol{w}_t, \tag{9}$$

where $\widetilde{\boldsymbol{\sigma}}(\cdot)$ is independent of $\boldsymbol{Z}_t$, but the parameters of (8) adn (9) are the same as in (7).

For notational convenience we will suppress the dependence of $\boldsymbol{\theta}$ and $\boldsymbol{u}_t$, i.e. we will use the notation

$$\boldsymbol{f}(\boldsymbol{X}_t, t) = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta}) \tag{10}$$

$$\boldsymbol{\sigma}(\boldsymbol{X}_t, t) = \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta}). \tag{11}$$

In real life systems $\boldsymbol{u}_t$ is often a set of observations, i.e. not a function that can be differentiated analytically, and this has to be kept in mind in the following development of the transformations.

3 One dimensional diffusion

The fundamental tool for transformations of SDE's is Itô's lemma (the version given below is due to Øksendal (2003))

Theorem 1 (Itô's lemma): *Let X_t be an Itô process given by*

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dw_t. \tag{12}$$

Let $\psi(X_t, t) \in C^2([0, \infty) \times \mathbb{R}$. Then

$$Z_t = \psi(X_t, t) \tag{13}$$

is again an Itô process, and

$$dZ_t = \frac{\partial \psi}{\partial t}(X_t, t)dt + \frac{\partial \psi}{\partial x}(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2}(X_t, t)(dX_t)^2, \tag{14}$$

where $(dX_t)^2$ is calculated according to the rules

$$dt \cdot dt = dt \cdot dw_t = dw_t \cdot dt = 0, \quad dw_t \cdot dw_t = dt. \tag{15}$$

The proof of this theorem is out of the scope of this note, and the reader is referred to Øksendal (2003).

It is illustrative to express Itô's formula in terms of dw_t rather than dX_t. For notational reasons we will sometimes write f for $f(X_t, t)$, σ for $\sigma(X_t, t)$ and ψ

for $\psi(X_t, t)$, partial derivatives will be written as $\psi_s = \frac{\partial \psi}{\partial s}$ and $\psi_{ss} = \frac{\partial^2 \psi}{\partial s^2}$. Rearranging (14) gives

$$dZ_t = \psi_t dt + \psi_x \cdot (f dt + \sigma dw_t) + \frac{1}{2}\psi_{xx} \cdot (f dt + \sigma dw_t)^2 \tag{16}$$

$$= (\psi_t + \psi_x \cdot f)\, dt + \psi_x \cdot \sigma dw_t + \frac{1}{2}\psi_{xx} \cdot \sigma^2 dt \tag{17}$$

$$= \left(\psi_t + \psi_x \cdot f + \frac{1}{2}\psi_{xx} \cdot \sigma^2\right) dt + \psi_x \cdot \sigma dw_t. \tag{18}$$

With this formulation we are ready for the construction of a transformation for removal of level dependent noise. The following constructive theorem is often referred to as the Lamperti transform (Iacus, 2008; Luschgy, 2006).

Theorem 2 (Lamperti transform): *Let X_t be an Itô process as in (12), and define*

$$\psi(X_t, t) = \int \frac{1}{\sigma(x,t)} dx \bigg|_{x=X_t}, \tag{19}$$

if ψ is one to one from the state space of X_t onto $\mathbb{R}$ for every $t \in [0, \infty)$, then choose $Z_t = \psi(X_t, t)$. Otherwise if $\sigma(X_t, t) > 0 \quad \forall (X_t, t)$ choose

$$Z_t = \psi(X_t, t) = \int_{\xi}^{x} \frac{1}{\sigma(u,t)} du \bigg|_{x=X_t}, \tag{20}$$

where ξ is some point inside the state space of X_t. Then Z_t has unit diffusion and is governed by the SDE

$$dZ_t = \Bigg(\psi_t(\psi^{-1}(Z_t,t),t) + \frac{f(\psi^{-1}(Z_t,t),t)}{\sigma(\psi^{-1}(Z_t,t),t)} - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t,t),t)\Bigg) dt + dw_t. \tag{21}$$

A transformation of the state-space clearly has to be one to one, such that every point in the state space of X_t can be uniquely identified by the inverse transformation of Z_t. If (19) is not one to one, then choosing the transformation (20) (due to Luschgy (2006)) will ensure that the transformation is one to one, since ψ is then a strictly increasing function of X_t. We will prove Eq. (19) and leave Eq. (20) to the reader.

PROOF. (Of Theorem 2) From (18) it is easy to realize that level dependent diffusion can be removed by choosing the transformation ψ as

$$\psi(X_t, t) = \int \frac{1}{\sigma(x,t)} dx \bigg|_{x=X_t} \implies \psi_x(X_t, t) = \frac{1}{\sigma(X_t, t)}. \tag{22}$$

Differentiation w.r.t. x and time gives

$$\psi_{xx}(X_t,t) = -\frac{\sigma_x(X_t,t)}{\sigma(X_t,t)^2} \tag{23}$$

$$\psi_t(X_t,t) = \frac{\partial}{\partial t}\int \frac{1}{\sigma(x,t)}dx\bigg|_{x=X_t}. \tag{24}$$

Inserting in (18) gives

$$dZ_t = \left(\frac{\partial}{\partial t}\int \frac{1}{\sigma(x,t)}dx\bigg|_{x=X_t} + \frac{f}{\sigma} - \frac{1}{2}\frac{\sigma_x}{\sigma^2}\sigma^2\right)dt + dw_t. \tag{25}$$

Cancelling out denominators and enumerators and inserting ψ_t and $X_t = \psi^{-1}(Z_t,t)$ gives the desired result. □

Theorem 2 gives a very useful approach for removal of level dependent noise. The discussion of the theorem in the following, is largely example driven and divided in two parts. 1) Time independent diffusion i.e. $\psi_t = 0$, and 2) time dependent diffusion.

3.1 Time independent diffusion

We begin this section with a small example, which illustrates the use of the Lamperti transformation.

Example 2 (Geometric Brownian motion): *Let X_t be an Itô process (SDE) given by*

$$dX_t = aX_t dt + \sigma X_t dw_t; \quad X_0 = 1, \tag{26}$$

where σ and a are real constants. Choose ψ as in (19), i.e.

$$Z_t = \psi(X_t) = \int \frac{1}{\sigma x}dx\bigg|_{x=X_t} = \frac{\log(X_t)}{\sigma}, \tag{27}$$

and

$$X_t = \psi^{-1}(Z_t) = e^{\sigma Z_t}. \tag{28}$$

By (21) Z_t is an Itô process given by

$$dZ_t = \left(\frac{aX_t}{\sigma X_t} - \frac{1}{2}\sigma \right) dt + dw_t \tag{29}$$

$$= \left(\frac{a}{\sigma} - \frac{1}{2}\sigma \right) dt + dw_t. \tag{30}$$

In this case the solution of Z_t can be given explicitly as

$$Z_t = \left(\frac{a}{\sigma} - \frac{1}{2}\sigma \right) t + w_t, \tag{31}$$

i.e. $Z_t \sim N\left(\left(\frac{a}{\sigma} - \frac{1}{2}\sigma\right)t, t\right) \Rightarrow \sigma Z_t \sim N\left(\left(a - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$, since $X_t = e^{\sigma Z_t}$, the solution of X_t is given as $X_t \sim LN\left(\left(a - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$, where LN is the log-normal distribution. □

In the example above the Lamperti transform actually solves the original equation. This is not the case in general, but Itô's formula can be used to solve SDE's, although the class of equations that are solvable in this fashion is limited. The inverse transform of Z_t was not a part of the SDE governing Z_t, this is not the case in general, and SDE's that are apparently very simple cannot be solve explicitly, as the next example illustrates.

Example 3 *Consider*

$$dX_t = (b + aX_t)dt + \sigma X_t dw_t. \tag{32}$$

Using (19) we get the same transformation as in Example 2, and Z_t is governed by

$$dZ_t = \left(\frac{b + aX_t}{\sigma X_t} - \frac{1}{2}\sigma \right) dt + dw_t \tag{33}$$

$$= \left(\frac{b}{\sigma} e^{-\sigma Z_t} + \frac{a}{\sigma} - \frac{1}{2}\sigma \right) dt + dw_t, \tag{34}$$

this SDE does not have an explicit solution, but parameter estimation is available through e.g. CTSM and numerical solutions (i.e. the distribution) can be found through simulations.

Eq. (19) is in principle always valid. The practical application of the transformation is, however, limited by our ability of find an explicit solution of the inverse transformation

$$X_t = \psi^{-1}(Z_t, t). \tag{35}$$

Such solutions are not always available as illustrated in the following example

Example 4 *Consider the diffusion process*

$$dX_t = f(X_t)dt + (\sigma_0 + \sigma_1\sqrt{X_t})dw_t. \tag{36}$$

The Lamperti transform becomes

$$Z_t = \psi(X_t) = \frac{\sigma_0}{\sigma_1^2}\log\left((\sigma_0 + \sigma_1\sqrt{X_t})^{-2}\right) + \frac{2}{\sigma_1}\sqrt{X_t} \tag{37}$$

$$= \frac{2}{\sigma_1}\left(\sqrt{X_t} - \frac{\sigma_0}{\sigma_1}\log\left(\sigma_0 + \sigma_1\sqrt{X_t}\right)\right). \tag{38}$$

In this case the Itô diffusion of Z_t cannot be written as an explicit function of Z_t, because X_t cannot be written as an explicit function of Z_t. □

As illustrated by Example 4 explicit solutions for the inverse transformation does not always exist, however many "real" life examples allow the explicit solution of the inverse transform. For instance explicit solutions of ψ^{-1} is available when $\sigma(X_t) = \sigma_1 X_t^{\gamma}$ for any constant γ, models of this type important in mathematical finance, where γ express the volatility of the market.

For models where $\sigma(X_t)$ are more complex, solutions to ψ^{-1} are in general not available. Biological models often use proportional or square root dependent diffusion terms, and in addition additive diffusion might be appropriate if the model contain additive input. As we saw in Example 4, ψ^{-1} is not available in this case. A quite flexible system where ψ^{-1} is available is the Pearson diffusion (Forman and Sørensen , 2008), which is considered in the following example.

Example 5 (Pearson diffusion): *Consider the diffusion process*

$$dX_t = f(X_t)dt + \sqrt{\sigma_0 + \sigma_1 X_t + \sigma_2 X_t^2}dw_t. \tag{39}$$

Actually this is an extension of the Pearson diffusion as the Pearson diffusion also have $f(X_t) = (b - aX_t)$. In this context we will, however, only consider the diffusion term. Use of the Lamperti transform (19) gives

$$Z_t = \psi(X_t) = \frac{1}{\sqrt{\sigma_2}}\log\left(\frac{\sigma_1}{2\sqrt{\sigma_2}} + \sqrt{\sigma_2}X_t + \sqrt{\sigma_0 + \sigma_1 X_t + \sigma_2 X_t^2}\right) \tag{40}$$

with the inverse

$$X_t = \psi^{-1}(Z_t) = \left(\frac{\sigma_1^2}{8\sigma_2^{3/2}} - \frac{\sigma_0}{2\sqrt{\sigma_2}}\right)e^{-\sqrt{\sigma_2}Z_t} + \frac{1}{2\sqrt{\sigma_2}}e^{\sqrt{\sigma_2}Z_t} - \frac{\sigma_1}{2\sigma_2} \tag{41}$$

and the Itô process for Z_t is given by

$$dZ_t = \left(\frac{f(\psi^{-1}(Z_t))}{\sqrt{\sigma_0 + \sigma_1 \psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} - \frac{\sigma_1 + 2\sigma_2 \psi^{-1}(Z_t)}{4\sqrt{\sigma_0 + \sigma_1 \psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} \right) dt + dw_t \tag{42}$$

$$= \frac{f(\psi^{-1}(Z_t)) - \frac{1}{4}\left(\sigma_1 + 2\sigma_2 \psi^{-1}(Z_t)\right)}{\sqrt{\sigma_0 + \sigma_1 \psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} dt + dw_t. \tag{43}$$

Clearly the resulting SDE is very complex, it will however provide the opportunity of testing hypothesis of $\sigma_i = 0$. In the construction of SDE's of the type discussed in this example it is important to ensure that the diffusion term exists for all X_t in the state space of X_t (we would need to examine the drift term at the boundary). □

Even though the Lamperti transform is limited by our ability of finding the inverse, it is still possible to use transformations that remove level dependent noise for quite general classes of diffusion processes, as illustrated in Example 5.

3.2 Time dependent diffusion

The SDE (21) depends on the time derivative of ψ, and even though this might be a quite complicated function, it is in principle always possible to find such a solution. In real life applications the time dependence of σ will, however, often be through an observed input, in this case the differentiation have to be done numerically. It might therefore be advisable to choose a transformation that leaves the diffusion term time dependent. This does however limit the the class of transformations substantially, it is e.g. not possible if one of the diffusion parameters in the Pearson diffusion depends on time.

In general it is possible to succeed in the case where the diffusion is given by

$$\sigma(X_t, t) = \alpha(t)\beta(X_t) \tag{44}$$

In this case use the Lamperti transform on $\beta(X_t)$ and leave the diffusion time-dependent, i.e. put

$$Z_t = \psi(X_t) = \int \frac{1}{\beta(x)} dx \bigg|_{x=X_t}, \tag{45}$$

and proceeding like in the time-independent diffusion we get

$$dZ_t = \left(\frac{f(\psi^{-1}(Z_t), t)}{\beta(\psi^{-1}(Z_t))} - \frac{1}{2}\beta_x(\psi^{-1}(Z_t))\alpha^2(t) \right) dt + \alpha(t) dw_t. \tag{46}$$

If the time dependence is either an explicit function of t or the differential of the time dependence is available through observations then Theorem 2 can still be applied, but the functional relationships do however become considerable more complex, as the next example illustrates.

Example 6 *Consider a process driven by a noisy time varying input $b(t)$ (birth process) and with a constant death-rate, the SDE formulation could be*

$$dX_t = (b(t) + aX_t)dt + (\sigma_0 b(t) + \sigma_1 X_t)dw_t, \tag{47}$$

where $a > 0$ and $b(t) > 0 \quad \forall t$. The Lamperti transform becomes

$$\psi(X_t, t) = \int \frac{1}{\sigma_0 b(t) + \sigma_1 x} dx \bigg|_{x=X_t} = \frac{\log(\sigma_0 b(t) + \sigma_1 x)}{\sigma_1} \tag{48}$$

implying

$$\psi_t(X_t, t) = \frac{\sigma_0 b'(t)}{\sigma_1(\sigma_0 b(t) + \sigma_1 X_t)} \tag{49}$$

$$\psi^{-1}(Z_t, t) = \frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1} \tag{50}$$

$$\sigma_x(X_t, t) = \sigma_1, \tag{51}$$

and $Z_t = \psi(X_t, t)$ is governed by the process

$$dZ_t = \left(\frac{\sigma_0 b'(t)}{\sigma_1(\sigma_0 b(t) + \sigma_1 \psi^{-1}(Z_t, t))} + \frac{b(t) + a\psi^{-1}(Z_t, t)}{\sigma_0 b(t) + \sigma_1 \psi^{-1}(Z_t, t)} - \frac{1}{2}\sigma_1 \right) dt + dw_t \tag{52}$$

$$= \left(\frac{\frac{\sigma_0}{\sigma_1} b'(t) + b(t) + a\frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1}}{\sigma_0 b(t) + \sigma_1 \frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1}} - \frac{1}{2}\sigma_1 \right) dt + dw_t \tag{53}$$

$$= \left[\left(\frac{\sigma_0}{\sigma_1} b'(t) + \left(1 - a\frac{\sigma_0}{\sigma_1}\right) b(t) + \frac{a}{\sigma_1} e^{\sigma_1 Z_t} \right) e^{-\sigma_1 Z_t} - \frac{1}{2}\sigma_1 \right) dt + dw_t$$

$$= \left\{ \left[\frac{\sigma_0}{\sigma_1} b'(t) + \left(1 - a\frac{\sigma_0}{\sigma_1}\right) b(t) \right] e^{-\sigma_1 Z_t} + \frac{a}{\sigma_1} - \frac{1}{2}\sigma_1 \right\} dt + dw_t. \tag{54}$$

In principle this is straight forward, but $b(t)$ will often be a function of some observed process and in this case we will therefore need observations of the differential of $b(t)$.

4 Multivariate Diffusion

The Lamperti transform presented so far is a univariate transformation, but it is possible to generalise this for a restricted class of multivariate diffusion processes. As for the one-dimensional diffusion process, Itô's lemma for multi-dimensional diffusion is the key to understand the multi-dimensional transformation. Again a good reference is Øksendal (2003).

Theorem 3 (Itô's lemma):

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t,)d\boldsymbol{w}_t, \tag{55}$$

with $t \in \mathbb{R}_+$ *being time,* $\boldsymbol{X}_t \in \mathbb{R}^n$ *the state vector,* $\boldsymbol{w}_t \in \mathbb{R}^n$ *multivariate standard Brownian motion,* $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ *and* $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$*. Then for a given transformation*

$$\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = [\psi_1(\boldsymbol{X}_t, t), ..., \psi_n(\boldsymbol{X}_t, t)], \tag{56}$$

where $\boldsymbol{\psi}$ *is a* C^2 *function from* $\mathbb{R}^n \times [0, \infty)$ *into* $\mathbb{R}^n$*,* $\boldsymbol{Z}_t$ *is again an Itô process given by*

$$\begin{aligned} dZ_{k,t} =& \frac{\partial \psi_k}{\partial t}(\boldsymbol{X}_t, t)dt + \sum_{i=1}^{n} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t, t)dX_{i,t} + \\ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(\boldsymbol{X}_t, t)dX_{j,t}dX_{i,t}. \end{aligned} \tag{57}$$

Where $dw_i dw_j = 0$ *for* $i \neq j$*,* $dw_i dw_i = dt$*, and* $dw_i dt = dt dw_i = dt dt = 0 \quad \forall i$*.*

In the version of Itô's lemma given above $\boldsymbol{\psi} \in \mathbb{R}^n$. In the general version of Itô's lemma this not a requirement, but we have restricted the attention to equal dimensions of $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$. The derivations below do however easily generalise.

It is again illustrative to write $dZ_{k,t}$ in terms of $dw_{i,t}$ rather than $dX_{i,t}$,

$$
\begin{aligned}
dZ_{k,t} =& \frac{\partial \psi_k}{\partial t}(\boldsymbol{X}_t, t)dt + \sum_{i=1}^{n} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t, t)dX_{i,t} + \\
& \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(\boldsymbol{X}_t, t)dX_{j,t}dX_{i,t} \qquad (58)\\
=& \left((\psi_k)_t + \sum_{i=1}^{n}(\psi_k)_{x_i} f_i \right) dt + \sum_{i=1}^{n}(\psi_k)_{x_i} \left(\sum_{h=1}^{n} \sigma_{ih} dw_{h,t} \right) + \\
& \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\psi_k)_{x_i,x_j} \left(\sum_{h=1}^{n} \sigma_{jh} dw_{h,t} \right) \left(\sum_{l=1}^{m} \sigma_{jl} dw_{l,t} \right) \qquad (59)\\
=& \left((\psi_k)_t + \sum_{i=1}^{n}(\psi_k)_{x_i} f_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\psi_k)_{x_i,x_j} \left(\sum_{h=1}^{m} \sigma_{jh}\sigma_{ih} \right) \right) dt + \\
& \sum_{h=1}^{m} \left(\sum_{i=1}^{n} (\psi_k)_{x_i} \sigma_{ih} \right) dw_{h,t}, \qquad (60)
\end{aligned}
$$

where subscript $\{h, i, j, k\}$ refer to elements of vectors and matrices, subscripts x_i and t refer to partial differentiation (except in $Z_{i,t}$ and $w_{i,t}$ where t refer to time). From the last expression in Eq. (60) it is seen that the removal of level dependent noise requires the solution of the following system of PDEs

$$
\sum_{i=1}^{n}(\psi_k)_{x_i}\sigma_{i1}(\boldsymbol{x}, t,) = c_1(t) \qquad (61)
$$

$$
\sum_{i=1}^{n}(\psi_k)_{x_i}\sigma_{i2}(\boldsymbol{x}, t) = c_2(t) \qquad (62)
$$

$$
\vdots
$$

$$
\sum_{i=1}^{n}(\psi_k)_{x_i}\sigma_{in}(\boldsymbol{x}, t) = c_n(t), \qquad (63)
$$

where c_i is an arbitrary function of t. Such a system can not be solved in general, since for given $\boldsymbol{\sigma}$, this results in n equations with one unknown (ψ_k).

Nielsen & Madsen (2001) claim that under the assumptions 1) $\sigma_{ij} \neq 0$ and 2)

$$
\sigma_{ij}(\boldsymbol{X}_t, t) = \sigma_{ij}(X_t^{\nu(i)}, t), \quad i = 1, ..., n, j = 1, ..., n, \qquad (64)
$$

it is possible to find a transformation. The application of Itô lemma is however wrong, we will not go through the proof of this, but applying (61)-(63) will lead to $\nu(i) = i$.

The difficulties of removing state dependent diffusion can be illustrated by a simple example.

Example 7 *Consider the diffusion process*

$$d \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} X_{2,t} & 0 \\ 0 & X_{1,t} \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}, \tag{65}$$

let $Z_{1,t} = \psi_1(\boldsymbol{X}_t)$ and $Z_{2,t} = \psi_2(\boldsymbol{X}_t)$. Using Itô's lemma we get

$$\begin{aligned} dZ_{1,t} =& \frac{\partial}{\partial x_1}\psi_1(X_{1,t}, X_{2,t})X_{2,t}dw_{1,t} + \frac{\partial}{\partial x_2}\psi_1(X_{1,t}, X_{2,t})X_{1,t}dw_{2,t} + \\ & \frac{1}{2}\left(\frac{\partial^2}{\partial x_1 \partial x_1}\psi_1(X_{1,t}, X_{2,t})X_{2,t}^2 + \frac{\partial^2}{\partial x_2 \partial x_2}\psi_1(X_{1,t}, X_{2,t})X_{1,t}^2\right)dt, \end{aligned} \tag{66}$$

the first term requires the solution of

$$c_1 = x_2 \frac{\partial}{\partial x_1}\psi_1(x_1, x_2) \tag{67}$$

implying

$$\psi_1(x_1, x_2) = c\frac{x_1}{x_2} + \tilde{\psi}_1(x_2). \tag{68}$$

where $\tilde{\psi}_1(x_2)$ is an arbitrary function of x_2. The second term therefore require the solution of

$$\begin{aligned} c_2 =& \frac{\partial}{\partial x_2}\left(c_1\frac{x_1}{x_2} + \tilde{\psi}_1(x_2)\right) & (69) \\ =& -c_1\frac{x_1}{x_2^2} + \frac{d}{dx_2}\tilde{\psi}_1(x_2), & (70) \end{aligned}$$

as $\tilde{\psi}_1(x_2)$ is not a function of x_1 this differential equation does not admit a solution.
□

Even if a general multivariate version of Theorem 2 is not available, it is possible to state the less general result

Theorem 4 (Multivariate Lamperti transform): *Let X_t be an Itô diffusion given by*

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t)\boldsymbol{R}(t)d\boldsymbol{w}_t, \tag{71}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n\times n}$ is any matrix function of t, and $\boldsymbol{\sigma}(\boldsymbol{X}_t, t) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal elements $\sigma_{i,i}(\boldsymbol{X}_t, t)$ given by

$$\sigma_{i,i}(\boldsymbol{X}_t, t) = \sigma_i(X_{i,t}, t). \tag{72}$$

Then the transformation

$$Z_{i,t} = \psi_i(X_{i,t}, t) = \int \frac{1}{\sigma_i(x,t)}dx\bigg|_{x=X_{i,t}}, \tag{73}$$

will result in an Itô process with the i'th element given by

$$dZ_{i,t} = \left(\frac{\partial}{\partial t}\psi_i(x,t)\big|_{x=\psi^{-1}(Z_{i,t},t)} + \frac{f_i(\boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t,t),t)}{\sigma_i(\psi_i^{-1}(Z_{i,t},t),t)} - \frac{1}{2}\frac{\partial}{\partial x}\sigma_i(\psi_i^{-1}(Z_{i,t},t),t)\right)dt + \sum_{j=1}^{n} r_{ij}(t)dw_{j,t}, \tag{74}$$

where $r_{ij}(t)$ are elements of $\boldsymbol{R}(t)$ and

$$\boldsymbol{X}_t = \boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t, t). \tag{75}$$

PROOF. Apply Theorem 2 to each $X_{i,t}$ □

The remarks about time dependent diffusion made in Section 3.2 also apply to the multidimensional case. A simple example illustrates the use of Theorem 4.

Example 8 (Two-dimensional Geometric Brownian motion): *Consider the process*

$$d\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix}dt + \begin{bmatrix} X_{1,t} & 0 \\ 0 & X_{2,t} \end{bmatrix}\begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix}d\boldsymbol{w}_t \tag{76}$$

$$= \boldsymbol{A}\boldsymbol{X}_t dt + \boldsymbol{\sigma}(\boldsymbol{X}_t)\boldsymbol{R}d\boldsymbol{w}_t, \tag{77}$$

with initial condition $\boldsymbol{X}_0 = 1$, choose $Z_{1,t} = \psi(\boldsymbol{X}_t) = \log(X_{1,t})$ and $Z_{2,t} = \psi(\boldsymbol{X}_t) = \log(X_{2,t})$, then

$$d\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix}dt + \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix}d\boldsymbol{w}_t, \tag{78}$$

with initial condition $\boldsymbol{Z}_0 = 0$ and the solution of (78) *is*

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix} t + \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} \boldsymbol{w}_t \tag{79}$$

$$=(\boldsymbol{a} - diag(\boldsymbol{R}\boldsymbol{R}^T))t + \boldsymbol{R}\boldsymbol{w}_t, \tag{80}$$

where $\boldsymbol{a} = diag(\boldsymbol{A})$is a vector with elements equal to the diagonal elements of $\boldsymbol{A}$. In this case $\boldsymbol{Z}_t$ follows a Gaussian distribution $\boldsymbol{Z}_t \sim N((\boldsymbol{a} - \frac{1}{2}diag(\boldsymbol{R}\boldsymbol{R}^T))t, \boldsymbol{R}\boldsymbol{R}^T t)$. $\boldsymbol{X}_t$ is therefore distributed according to a multivariate $\log$-normal distribution, with the same parameters. □

The transformation presented in Theorem 4 is not a true Lamperti transform, since it does not transform to unit diffusion. This can be solved by the following theorem

Theorem 5 (Tranformation to unit diffusion): *Let X_t be an Itô diffusion given by*

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{R}(t)d\boldsymbol{w}_t, \tag{81}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n \times n}$ is any invertible matrix function of t, then the transformation

$$\boldsymbol{Z}_{i,t} = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = \boldsymbol{R}(t)^{-1}\boldsymbol{X}_t \tag{82}$$

will result in an Itô process given by

$$d\boldsymbol{Z}_t = \left[\left(\frac{d}{dt} \boldsymbol{R}(t)^{-1} \right) \boldsymbol{R}(t)\boldsymbol{Z}_t + \boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{R}(t)\boldsymbol{Z}_t, t) \right] dt + d\boldsymbol{w}_t \tag{83}$$

(84)

with $(\frac{d}{dt}\boldsymbol{R}(t)^{-1})$ the elements-wise derivative of $\boldsymbol{R}(t)^{-1}$ and

$$\boldsymbol{X}_t = \boldsymbol{R}(t)\boldsymbol{Z}_t. \tag{85}$$

PROOF. Consider the i'th coordinate of $d\boldsymbol{X}_t$

$$dX_{i,t} = f_i(\boldsymbol{X}_t, t)dt + \sum_{j=1}^{n} (\boldsymbol{R}(t))_{i,j} dw_j, \tag{86}$$

and the i'th coordinate of $\boldsymbol{Z}_t$

$$Z_{i,t} = \psi_i(\boldsymbol{X}_t, t) = \sum_{j=1}^{n} (\boldsymbol{R}(t)^{-1})_{ij} X_{j,t} \tag{87}$$

by Itô's lemma we get (noting that $\frac{\partial^2}{\partial x_i \partial x_j}\psi_i(\boldsymbol{x},t) = 0 \; \forall \; i \in \{1,..,n\}$, and $j \in \{1,...,n\}$)

$$Z_{i,t} = \frac{\partial}{\partial t}\psi_i(\boldsymbol{X}_t,t)dt + \sum_{j=1}^{n}\frac{\partial}{\partial x_j}\psi_i(\boldsymbol{X}_t,t)dX_{j,t} \tag{88}$$

$$= \frac{\partial}{\partial t}\sum_{j=1}^{n}(\boldsymbol{R}(t)^{-1})_{ij}X_{j,t}dt + \sum_{j=1}^{n}(\boldsymbol{R}(t)^{-1})_{ij}[f_j(\boldsymbol{X}_t,t)dt + \sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}dw_h] \tag{89}$$

$$= \left[\sum_{j=1}^{n}\frac{d}{dt}\boldsymbol{R}(t)_{ij}^{-1}X_{j,t} + \sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}f_j(\boldsymbol{X}_t,t)\right]dt + \sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}\sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}dw_h \tag{90}$$

$$= \left[\sum_{j=1}^{n}\left(\frac{d}{dt}\boldsymbol{R}(t)_{ij}^{-1}\right)\sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}^{-1}Z_{h,t} + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt + \sum_{h=1}^{n}\left(\sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}\boldsymbol{R}(t)_{jh}\right)dw_h \tag{91}$$

$$= \left[\sum_{h=1}^{n}\left(\left(\frac{d}{dt}\boldsymbol{R}(t)^{-1}\right)\boldsymbol{R}(t)\right)_{ih}Z_{h,t} + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt \sum_{h=1}^{n}\left(\boldsymbol{R}(t)^{-1}\boldsymbol{R}(t)\right)_{ih}dw_h \tag{92}$$

$$= \left[\left(\left(\frac{d}{dt}\boldsymbol{R}(t)^{-1}\right)\boldsymbol{R}(t)\boldsymbol{Z}_t\right)_i + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt + dw_i, \tag{93}$$

writing the matrix formulation of the above gives the desired result. □

Combining Theorem 4 and 5 gives a multivariate version of the Lamperti transform. This is illustrated by applying Theorem 5 to Example 8.

Example 9 *Consider the transformed process of Example 8, then*

$$\boldsymbol{R}^{-1} = \frac{1}{\det(\boldsymbol{R})}\begin{bmatrix} r_{22} & -r_{12} \\ -r_{21} & r_{11} \end{bmatrix}, \tag{94}$$

and the process

$$\tilde{\boldsymbol{Z}}_t = \tilde{\psi}(\boldsymbol{Z}_t) = \boldsymbol{R}^{-1}\boldsymbol{Z}_t, \tag{95}$$

is given by

$$d\begin{bmatrix} \tilde{Z}_{1,t} \\ \tilde{Z}_{2,t} \end{bmatrix} = \frac{1}{\det(\boldsymbol{R})}\begin{bmatrix} r_{2,2} & -r_{1,2} \\ -r_{2,1} & r_{1,1} \end{bmatrix}\begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix} dt + \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix} \tag{96}$$

and $\tilde{\boldsymbol{Z}}_t \sim N(\boldsymbol{R}^{-1}(\boldsymbol{a} - \frac{1}{2}diag(\boldsymbol{R}\boldsymbol{R}^T))t, \boldsymbol{I}t)$ *and the inverse of* $\tilde{\boldsymbol{Z}}$ *is*

$$\begin{aligned} \boldsymbol{X}_t =& \boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t) = \boldsymbol{\psi}^{-1}(\tilde{\boldsymbol{\psi}}^{-1}(\tilde{\boldsymbol{Z}}_t)) = \boldsymbol{\psi}^{-1}(\boldsymbol{R}\tilde{\boldsymbol{Z}}_t)) = \exp(\boldsymbol{R}\tilde{\boldsymbol{Z}}_t) \\ =& \begin{bmatrix} e^{r_{11}\tilde{Z}_{1,t} + r_{12}\tilde{Z}_{2,t}} \\ e^{r_{21}\tilde{Z}_{1,t} + r_{22}\tilde{Z}_{2,t}} \end{bmatrix}. \end{aligned} \tag{97}$$

□

Theorem 5 states that a process with state independent diffusion can be written as a weighted version of the original process, which has unit diffusion. The weight is the inverse of diffusion matrix, it is tempting to interpret the diffusion matrix ($\boldsymbol{R}(t)$) as a local standard deviation, such an interpretation is however not straight forward, since the local variance is different from $\boldsymbol{R}(t)^2$. The density of an SDE is determined by the Fokker-Planck equation (sometimes referred to as the Kolmogorov forward equation), which in the multidimensional case is given by (Gard, 1988)

$$\begin{aligned} p_t(\boldsymbol{x}, t) =& -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right) + \\ & \frac{1}{2}\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j}\sum_k \left(\sigma_{ik}(\boldsymbol{x}, t)\sigma_{jk}(\boldsymbol{x}, t)\right) p(\boldsymbol{x}, t) \end{aligned} \tag{98}$$

$$\begin{aligned} =& -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right) + \\ & \frac{1}{2}\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j}\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^T(\boldsymbol{x}, t)\right)_{ij} p(\boldsymbol{x}, t). \end{aligned} \tag{99}$$

The density ($p(\cdot)$) does not depend on $\boldsymbol{\sigma}$ itself, but only on $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$, meaning that p_t is only uniquely determined up to what can loosely be refereed to as the (non unique) "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. However for a positive definite symmetric matrix, say $\boldsymbol{A}$, there exist a unique positive definite symmetric matrix $\boldsymbol{T}$, such that $\boldsymbol{T}^2 = \boldsymbol{A}$, and by construction $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is a positive definite symmetric matrix if $\boldsymbol{\sigma}$ has full rank. Again the best way to understand this is by considering a small example.

Example 10 *Consider the SDE*

$$d\boldsymbol{X}_t = \begin{bmatrix} 2 & -3 \\ -5 & -4 \end{bmatrix} d\boldsymbol{w}_t; \quad \boldsymbol{X}_0 = \boldsymbol{0} \tag{100}$$

it is well known that $\boldsymbol{X}_t$ follow a Gaussian distribution with mean $\boldsymbol{0}$, the variance for this process is given by

$$\begin{aligned} V(\boldsymbol{X}_t) =& \boldsymbol{\sigma}\boldsymbol{\sigma}^T t & (101) \\ =& \begin{bmatrix} 2 & -3 \\ -5 & -4 \end{bmatrix} \begin{bmatrix} 2 & -5 \\ -3 & -4 \end{bmatrix} t & (102) \\ =& \begin{bmatrix} 13 & 2 \\ 2 & 41 \end{bmatrix} t. & (103) \end{aligned}$$

Consider now the proces

$$d\tilde{\boldsymbol{X}}_t = \begin{bmatrix} 3.6 & 0.2 \\ 0.2 & 6.4 \end{bmatrix} d\boldsymbol{w}_t; \quad \boldsymbol{X}_0 = \boldsymbol{0}, \tag{104}$$

the variance of $\tilde{\boldsymbol{X}}_t$ is

$$\begin{aligned} V(\tilde{\boldsymbol{X}_t}) =& \tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^T t & (105) \\ =& \begin{bmatrix} 3.6 & 0.2 \\ 0.2 & 6.4 \end{bmatrix}^2 t & (106) \\ =& \begin{bmatrix} 3.6^2 + 0.2^2 & 0.2(6.4 + 3.6) \\ 0.2(6.4 + 3.6) & 6.4^2 + 0.2^2 \end{bmatrix} t & (107) \\ =& \begin{bmatrix} 13 & 2 \\ 2 & 41 \end{bmatrix} t. & (108) \end{aligned}$$

Which shows that $\tilde{\boldsymbol{X}}_t$ is a weak solution to the SDE (100).

Analytic solutions for the unique "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ are not easy to find. This is however not important if we are interested in estimation, since the likelihood is generated by the weak solution to the SDE. The important conclussion is that we can only indetify the number of parameters corresponding to a symetric version of the "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$.

In the one dimensional case we can think $\boldsymbol{R}$ as standard deviation is the following sense, let x_t be a continuous time random walk

$$dx_t = r dw_t, \tag{109}$$

then the discrete time stochastic process (with $t_{k-1} < t_k$)

$$y_k = \frac{x_{t_k} - x_{t_{k-1}}}{r\sqrt{t_k - t_{k-1}}}; \quad k = 1, 2, \dots \quad , \tag{110}$$

is a sequence of iid. standard Gaussian random variables. The multidimensional equivalent to (109) is

$$d\boldsymbol{x}_t = \boldsymbol{R} d\boldsymbol{w}_t. \tag{111}$$

Now if the matrix $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ is symmetric and positive definite then $\boldsymbol{R}^2 = \boldsymbol{R}\boldsymbol{R}^T$, and the discrete time stochastic process (again with $t_{k-1} < t_k$)

$$\boldsymbol{y}_k = \frac{1}{\sqrt{t_k - t_{k-1}}} \boldsymbol{R}^{-1}(\boldsymbol{x}_{t_k} - \boldsymbol{x}_{t_{k-1}}); \quad k = 1, 2, \dots \quad , \tag{112}$$

is a sequence of n-dimensional iid. standard Gaussian random variables, and in this sense we can think of $\boldsymbol{R}$ as the standard deviation. The generation of correlated random variable is often done by simulating independent standard random numbers and then multiplying by the covariation matrix (Madsen, 2008). The transformation in Theorem 5 can be viewed as the SDE equivalent to such a transformation.

The examples presented so far have been rather simple with the purpose to explain or clarify the theory, for such purposes it is not illustrative to include more physical reasoning. It might however be motivating to see an example based on real life reasoning, the last example of this note is such an example.

Example 11 A competition model: *Consider a controlled experiment with two living populations P_1 and P_2 (e.g. bacteria or phytoplankton) eating the same two nutrients N_1 and N_2 (e.g. nitrogen and phosphor), but not each other. Let the experiment be constructed such the the total amount of nutrients are held constant. Biological growth models are often assumed to follow Liebigs law of minimum and Michaelis-Menten kinetics, i.e.*

$$dP_i = \left(\min \left(\frac{\mu_{i,1} N_1}{k_{i,1} + N_1}, \frac{\mu_{i,2} N_2}{k_{i,2} + N_2} \right) - m_i \right) P_i dt \tag{113}$$

$$= (f_i(\boldsymbol{N}) - m_i) P_i dt, \tag{114}$$

where $m_i > 0$ is the mortality rate and $\min(\cdot)$ express the limiting factor. Further let a_{ij} be conversion factors that convert population i to nutrient j, such factors *are often known or approximately known from literature. As discussed earlier the diffusion term for biological processes is often assumed to be proportional to either* P_i *of* $\sqrt{P_i}$*, here we assume that the diffusion is proportional to* $P_i^{\gamma_i}$ *with* $\gamma_i \in (\frac{1}{2}, 1)$ *and leave it to the estimation procedure to determine* γ_i*.*

The SDE for the system described above is

$$d\begin{bmatrix} N_{1,t} \\ N_{2,t} \\ P_{1,t} \\ P_{2,t} \end{bmatrix} = \begin{bmatrix} -a_{11}(f_1(\boldsymbol{N})-m_1) & -a_{12}(f_2(\boldsymbol{N})-m_2) \\ -a_{21}(f_1(\boldsymbol{N})-m_1) & -a_{22}(f_2(\boldsymbol{N})-m_2) \\ f_1(\boldsymbol{N})-m_1 & 0 \\ 0 & f_2(\boldsymbol{N})-m_2 \end{bmatrix} \begin{bmatrix} P_{1,t} \\ P_{2,t} \end{bmatrix} dt + \begin{bmatrix} -a_{11}\sigma_1 P_{1,t}^{\gamma_1} & -a_{12}\sigma_2 P_{2,t}^{\gamma_2} \\ -a_{21}\sigma_1 P_{1,t}^{\gamma_1} & -a_{22}\sigma_2 P_{2,t}^{\gamma_2} \\ \sigma_1 P_{1,t}^{\gamma_1} & 0 \\ 0 & \sigma_2 P_{2,t}^{\gamma_2} \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}. \tag{115}$$

Seemingly we cannot apply the derived methods to transform this system to a system with constant diffusion, however the above system have a 2-dimensional distribution only and transformation is therefore possible. Using Theorem 4 on P_i *gives*

$$\tilde{P}_{i,t} = \psi_i(P_{i,t}) = \frac{1}{\sigma_i}\int x_i^{-\gamma_i}dx\bigg|_{x=P_{i,t}} = \frac{P_{i,t}^{1-\gamma_i}}{\sigma_i(1-\gamma_i)}, \tag{116}$$

with the inverse function given by

$$P_{i,t} = (1-\gamma_i)^{\gamma_i-1}(\sigma_i\tilde{P}_{i,t})^{\gamma_i-1} = (\tilde{\gamma}_i\sigma_i\tilde{P}_{i,t})^{-\tilde{\gamma}_i}, \tag{117}$$

where $\tilde{\gamma}_i = 1-\gamma_i$*. Now choose*

$$\tilde{N}_{i,t} = \phi_i(N_{i,t}, P_{1,t}, P_{2,t}) = N_{i,t} + a_{i1}P_{1,t} + a_{i2}P_{1,t}. \tag{118}$$

Using Itô's formula we get

$$\begin{aligned} d\tilde{N}_{i,t} =& (-a_{i1}(f_1(\boldsymbol{N})-m_1)P_{1,t} - a_{i2}(f_2(\boldsymbol{N})-m_2)P_{2,t})dt \\ &- a_{i1}\sigma_1 P_{1,t}dw_{1,t} - a_{i2}\sigma_2 P_{2,t}dw_{2,t} \\ &+ a_{i1}(f_1(\boldsymbol{N})-m_1)P_{1,t}dt + a_{i1}\sigma_1 P_{1,t}dw_{1,t} \\ &+ a_{i2}(f_2(\boldsymbol{N})-m_2)P_{2,t}dt + a_{i2}\sigma_2 P_{2,t}dw_{2,t} \qquad (119) \\ =& 0. \qquad (120) \end{aligned}$$

Implying that $\tilde{N}_{i,t} = \tilde{N}_{i,0}$ is constant and $N_{i,t}$ is given by

$$N_{i,t} = N_i(\boldsymbol{P}_t) \tag{121}$$

$$= \tilde{N}_{i,t} - a_{i1}P_{1,t} - a_{i2}P_{2,t} \tag{122}$$

$$= \tilde{N}_{i,0} - a_{i1}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{i2}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}. \tag{123}$$

The system equation for the transformed system takes the form

$$d\begin{bmatrix} \tilde{P}_{1,t} \\ \tilde{P}_{2,t} \end{bmatrix} = \begin{bmatrix} (\sigma_1\tilde{\gamma}_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1}(\tilde{f}_1(\tilde{\boldsymbol{P}}) - m_1) \\ (\sigma_2\tilde{\gamma}_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}(\tilde{f}_2(\tilde{\boldsymbol{P}}) - m_2) \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}, \tag{124}$$

with

$$\tilde{f}_i(\tilde{\boldsymbol{P}}) = \min\left(\frac{\mu_{i,1}N_1(\tilde{\boldsymbol{P}})}{k_{i,1} + N_1(\tilde{\boldsymbol{P}})}, \frac{\mu_{i,2}N_2(\tilde{\boldsymbol{P}})}{k_{i,2} + N_2(\tilde{\boldsymbol{P}})}\right) \tag{125}$$

$$= \min\left(\frac{\mu_{i,1}(\tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2})}{k_{i,1} + \tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}}, \frac{\mu_{i,2}(\tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2})}{k_{i,2} + \tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}}\right). \tag{126}$$

The derivations above strongly depend on the fact that the actual dimension of the joint distribution at time t is only 2, if there had been a random input of nutrient to the system, the derivation would not have been possible. It is therefore a crucial assumption that the experiment is conducted in a controlled environment, with no random interactions with the surroundings.

For the sake of completeness we will give an example of the observation equation, where we will assume that we are able to observe all the states of the original system, and that these observations are $\log$-normally distributed around the true state, i.e.

$$\begin{bmatrix} Y_{N_1,k} \\ Y_{N_2,k} \\ Y_{P_1,k} \\ Y_{P_2,t_k} \end{bmatrix} = \begin{bmatrix} \log(N_{1,t_k}) \\ \log(N_{2,t_k}) \\ \log(P_{1,t_k}) \\ \log(P_{2,t_k}) \end{bmatrix} + \begin{bmatrix} \epsilon_{N_1,t} \\ \epsilon_{N_2,t} \\ \epsilon_{P_1,t} \\ \epsilon_{P_2,t} \end{bmatrix} \tag{127}$$

$$= \begin{bmatrix} \log(\tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t_k})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t_k})^{-\tilde{\gamma}_2}) \\ \log(\tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t_k})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t_k})^{-\tilde{\gamma}_2}) \\ -\tilde{\gamma}_1(\log(\tilde{\gamma}_1) + \log(\sigma_1) + \log(\tilde{P}_{1,t_k})) \\ -\tilde{\gamma}_2(\log(\tilde{\gamma}_2) + \log(\sigma_2) + \log(\tilde{P}_{2,t_k})) \end{bmatrix} + \boldsymbol{\epsilon}_t, \tag{128}$$

where $\boldsymbol{\epsilon}_{t_k}$ follow a Gaussian distribution with mean zero variance $\boldsymbol{S}$. □

5 Summary and conclusion

We have shown how a class SDE's with state dependent diffusion can be transformed into SDE's with state independent diffusion. For one dimensional systems this transformation is rather straight forward and is only limited by the ability to find a closed form inverse transformation. Such transformations are important both in estimation and simulations. Iacus (2008) notes that the Lamperti transformation or similar transformations (not necessarily to unity) should always be used before simulation and that many estimation techniques rely on unit or constant diffusion. Luschgy (2006) presents proofs of convergence rates for a simulations procedure, which also relies on the existence of the Lamperti transform.

For time dependent diffusion the transformed process will depend on the time derivative of the transformation, which is equivalent to dependence on the time-derivative of the diffusion term. While this might be reasonable when the functional relation between the diffusion and time is given in an explicit form, it is problematic if the time dependence on the diffusion is through an observed input, because numerical differentiation will be needed.

For multidimensional diffusion processes the transformation to systems with state independent diffusion is more delicate, and Luschgy (2006) note that the Lamperti transform is essentially a one-dimensional transformation. This is also what has been shown here, however, it is also stressed that even with the restriction given in Theorem 4, there is still a large class of SDE's that can be handled through transformations. This class includes processes that seemingly is not included in Theorem 4, like the mass balance model presented in Example 11.

It is shown that the transformation to unit diffusion can be interpreted as a weighting with the local standard deviation. This means that the system noise innovation is equal for all states in the transformed process.

References

Aït-Sahalia Y. (2008) Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, **32**, (2), 906-937

Forman, J., and Sørensen, M. (2008) The Pearson Diffusion: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statics*, **35**, 438-465

Gard T.C. (1988) Introduction to stochastic differential equations. *Marcel Dekker, Inc.*, New York.

Iacus S. M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples. *Springer Series in Statistics*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory. *Academic Press*, New York.

Klebaner F. C. (2005) Introduction to Stochastic Calculus with Applications., *Imperial College Press*

Kristensen N.R., and Madsen H. (2003) Continuous time stochastic modeling - CTSM 2.3 - Mathematics Guide. *Technical University of Denmark*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models. *Automatica*, **40**, 225-237

Luschgy H. & Pagés G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications* **116**, *310-336*

Madsen, H. (2008) *Time Series Analysis.* Chapman & Hall/CRC

Nielsen J. N. and Madsen H. (2001) Applying the EKF to stochastic differential equations with level effects. *Automatica*, **37**, *107-112*

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer, Berlin*