Approximation algorithm supported on minimizing the Kullback-Leibler information divergence in some class of dynamical systems (October 2015)

Algorytm aproksymacyjny w oparciu o informację Kullbacka-Leiblera w pewnej klasie systemów dynamicznych

Jan Owedyk[1], Zdzisław Mathia[2], Hubert Zarzycki[3]

Treść: W pracy przedstawiono algorytm, który umożliwia skonstruowanie przybliżonych rozwiązań dla pewnej klasy systemów dynamiczych opisujących ewolucję w czasie gęstości prawdopodobieństwa. Przybliżone rozwiązania otrzymujemy minimalizując informację Kullbacka-Leiblera przy dodatkowych warunkach.
Wykazano, że pochodna informacji Kullbacka-Leiblera dla dokładnych i przybliżonych rozwiązań jest opisana przez tą samą formułę. W konsekwencji gdy w dynamicznym systemie maleje informacja Kullbacka-Leiblera dla dokładnych rozwiązań to także maleje dla przybliżonych rozwiązań.

Słowa kluczowe: Algorytm aproksymacyjny, Informacja Kullbacka-Leiblera, Metoda minimalizacji, Równania Fokkera-Plancka

Abstract: In this work an algorithm is presented for creating approximate solutions in some class of dynamical systems describing the time evolution probability densities. The approximate solutions are obtained by minimizing Kullback-Leibler divergence under some constrains.
It is shown that the derivatives of the Kullback-Leibler divergence for exact solutions and for approximate solutions are described by the same formula. In consequence if in a dynamical system the Kullback-Leibler divergence decreases in time for exact solutions, it also decreases for approximate solutions.

Keywords: Approximation algorithm, Kullback-Leibler divergence, Minimization methods, Fokker-Planck equation

I. Introduction

In the early eighties of the last century the Kullback-Leibler divergence was used to obtain approximate solutions in dynamical systems represented by n-dimentional Fokker-Planck equation [4] also with time dependent drift and diffusion coefficients [6] and in systems represented by master equations [5,7].
In this paper this approach is generalised for some dynamical systems in that the Kullback-Leibler information divergence $K(P,P_0) = \int P\, ln(P/P_0)\, d\boldsymbol{x} \geq 0,$ [2,3] satisfies inequality $dK/dt \leq 0$ for any two probability density functions $P = P(\boldsymbol{x},t)$, $P_0 = P_0(\boldsymbol{x},t)$ of continuous random variable defined on $\boldsymbol{x} = (x_1,\dots, x_n)$, which describe time evolution in the system. We have formulated a criterion on choosing such a dynamical system which will be called the Kullback-Leibler system.
In the dynamical system an approximate solution as an exponential probability density is postulated

$$P^{*}=P_0\, exp(-\lambda_0(t)-\sum_{i=1}^{N}\lambda_i(t)\, f_i)$$

obtained by minimizing the Kullback-Leibler divergence under some constrains. The functions $\lambda_i(t)$, $i=0,1,\dots,N$ are the Lagrange multipliers and their time evolution is determined by some system of ordinary differential equations and $f_i = f_i(\boldsymbol{x})$, $i = 1,\dots,N$ are some lineary independent functions. The probability density $P_0 = P_0(\boldsymbol{x},t)$ is a fixed exact solution of the Kullback-Leibler system.
This paper is organised as follows. In section 2 a definition of the Kullback-Leibler system is formulated and the stability of exact solutions in this system is investigated. In section 3 the approximate solution is obtained for the Kullback-Leibler system by minimizing Kullback-Leibler divergence under some conditions. In section 4 the stability of the approximate solutions is investigated. In section 5 the accompanied differential evolution equations for $\lambda_i(t)$, $i = 1,\dots,N$ are derived and investigated. Two important properties of solutions to this accompanied differential evolution equations are formulated. An optimization algorithm supported on minimizing the Kullback-Leibler divergence in the Kullback-Leibler systems is also formulated.

1. Department of Informatics, Kujawy and Pomorze University, Bydgoszcz, Poland j.owedyk@kpsw.edu.pl
2. Department of Informatics, Kujawy and Pomorze University, Bydgoszcz, Poland
3. Faculty of Computer Science, Wroclaw School of Information Technology, ul. Wejherowska 28,54-239 Wrocław, Poland, hzarzycki@horyzont.eu

II. The Kullback-Leibler Systems

We take for our considerations some dynamical systems describing by the equation

$$\partial P/\partial t = S_t P, \qquad (2.1)$$

where $P = P(\mathbf{x},t)$ is the time dependent probability density function of the continuous random variable defined on $\mathbf{x} = (x_1, ..., x_n)$ $\mathbf{E}^n$ ($\mathbf{E}^n$ *is the n-dimensional Euclidean space),* S_t is a tin ∊ dependent system operator. In the case of the Fokker-Planck equation, S_t is given by equation (A.2) (see appendix).

For ensuring the normalization condition $\int P(\mathbf{x},t)\, d\mathbf{x} = 1$ it must be satisfied

$$\int S_t P\, d\mathbf{x} = 0 \qquad (2.2)$$

for any probability density function $P(\mathbf{x},t)$.

We use the Kullback-Leibler divergence [2,3]

$$K(P,P_0) = \int P \ln(P/P_0)\, d\mathbf{x} \geq 0 \qquad (2.3)$$

as "a measure of distance" between any two solutions $P = P(\mathbf{x},t)$, $P_0 = P_0(\mathbf{x},t)$ of equation (2.1).

In order to investigate time dependence of Kullback-Leibler divergence we calculate its time derivative

$$dK/dt = \int ((\partial P/\partial t)\ln(P/P_0) + P\partial(\ln(P/P_0))/\partial t)d\mathbf{x} = \int ((\partial P/\partial t)\ln(P/P_0) + (\partial P/\partial t) - (\partial P_0/\partial t)(P/P_0))d\mathbf{x} = \int((S_t P)\ln(P/P_0) - (S_t P_0)(P/P_0))d\mathbf{x}. \qquad (2.4)$$

Above we have used normalisation to the unity i.e. $\int P\, d\mathbf{x} = 1$ and equations $\partial P/\partial t = S_t P$, $\partial P_0/\partial t = S_t P_0$ (P and P_0 satisfy (2.1)).

We will restrict ourselves to the system operator S_t so that the last formula in (2.4) is negative for any P and P_0 i.e.

$$\int((S_t P)\ln(P/P_0) - (S_t P_0)(P/P_0))d\mathbf{x} \leq 0, \ (=0 \text{ only if } P = P_0). \qquad (2.5)$$

The system described by the system operator S_t which satisfies inequality (2.5) will be called Kullback-Leibler system and the system operator S_t will be called Kullback-Leibler system operator. It is shown in Appendix that systems described by the Fokker-Planck equation are the Kullback-Leibler systems. From now we will consider only Kullback-Leibler systems and Kullback-Leibler system operators S_t.

According to the inequality (2.5) it follows that

$$dK/dt \leq 0, \ (=0 \text{ only if } P = P_0). \qquad (2.6)$$

The inequality (2.6) may be treated as a generalised *H*-theorem [6]. We can see that the inequality (2.5) is a criterion of choosing dynamical systems in which the generalised *H*-theorem is satisfied.

Using the above inequality one can investigate an asymptotic behaviour of solutions of the equation (2.1). Because the Kullback-Leibler divergence $K(P,P_0)$ (see (2.3)) is bounded from below and (2.6) is satisfied, one can write

$$\lim_{t\to\infty} dK/dt = 0.$$

If additionally from (2.7) and (2.5) it follows that

$$\lim_{t\to\infty} (P(\mathbf{x},t)/P_0(\mathbf{x},t)) = 1, \text{ for every } \mathbf{x}, \qquad (2.8)$$

then because the probability densities are normed to the unity, the difference between two arbitrary solutions $P = P(\mathbf{x},t)$, $P_0 = P_0(\mathbf{x},t)$ of equation (2.1) vanishes as time goes to infinity, i.e.

$$\lim_{t\to\infty} (P(\mathbf{x},t) - P_0(\mathbf{x},t)) = 0, \text{ for every } \mathbf{x}. \qquad (2.9)$$

III. The Minimizing Kullback-Leibler Divergence Solutions

Let $P_0(\mathbf{x},t)$ be some fixed solution of equation (2.1). An arbitrary solution $P(\mathbf{x},t)$ of the equation (2.1) may be written in the form

$$P = P(\mathbf{x},t) = P_0(\mathbf{x},t) \exp(-F(\mathbf{x},t)), \qquad (3.1)$$

where $F = F(\mathbf{x},t)$ is some function.

From (3.1) and (2.1) we obtain

$$\partial P/\partial t = (\partial P_0/\partial t) \exp(-F) - (\partial F/\partial t) P_0 \exp(-F) \qquad (3.2)$$

and

$$(\partial F/\partial t) P = (S_t P_0) \exp(-F) - S_t P. \qquad (3.3)$$

From (3.3) and (3.1) we have

$$(\partial F/\partial t) = (S_t P_0)/P_0 - (S_t P)/P \qquad (3.4)$$

and

$$(\partial F/\partial t) = (S_t P_0)/P_0 - (S_t (P_0 \exp(-F))) / (P_0 \exp(-F)). \qquad (3.5)$$

Equation (3.5) determines time evolution of the function $F = F(\mathbf{x},t)$.

We assume that the datas of the system are represented by mean values

$$\langle f_i \rangle_P = \int f_i(\mathbf{x}) P(\mathbf{x},t)\, d\mathbf{x}, \ i = 1,...,N \qquad (3.6)$$

of N linearly independent (together with $f_0 = f_0(\mathbf{x}) = 1$) functions $f_i = f_i(\mathbf{x})$, $i = 1,...,N$. According to the (3.6), (2.1) and (3.1) the evolution equations for the mean values are in

the following form

$$d\langle f_i \rangle_P/dt = \int f_i\, \partial P/\partial t\, d\mathbf{x} = \int f_i S_t P\, d\mathbf{x} = \int f_i S_t (P_0 \exp(-F))\, d\mathbf{x}, \ i = 1,..., N, \qquad (3.7)$$

finally

$$d\langle f_i \rangle_P/dt = \int f_i S_t (P_0 \exp(-F))\, d\mathbf{x}, \ i = 1,..., N. \qquad (3.8)$$

Using (3.3) and (3.1) we obtain useful formulas

$$\langle f_i \partial F/\partial t \rangle_p = \int f_i\, \partial F/\partial t\, P\, d\mathbf{x} = \int f_i (S_t P_0) \exp(-F)\, d\mathbf{x} - \int f_i S_t (P_0 \exp(-F))\, d\mathbf{x}, \ i = 0,1,...,N. \qquad (3.9)$$

For $i = 0$ we remember that $f_0 = f_0(\mathbf{x}) = 1$ then from (3.9) we have

$$\langle \partial F/\partial t\rangle_p = \int (S_t P_0)\, exp(-F)\, d\mathbf{x} - \int S_t (P_0 exp(-F))\, d\mathbf{x}. \tag{3.10}$$

From (3.1), (2.1) and normalization condition $\int P(\mathbf{x},t)\, d\mathbf{x} = 1$ we have

$$\int S_t (P_0 exp(-F))\, d\mathbf{x} = \int S_t P\, d\mathbf{x} = \int \partial P/\partial t\, d\mathbf{x} = d(\int P d\mathbf{x})/dt = 0, \tag{3.11}$$

then

$$\langle \partial F/\partial t\rangle_p = \int (S_t P_0)\, exp(-F)\, d\mathbf{x}. \tag{3.12}$$

In general the evolution equations (3.8) are not closed with respect to the mean values $\langle f_i\rangle_P = \int f_i(\mathbf{x}) P(\mathbf{x},t)\, d\mathbf{x}$ of functions $f_i = f_i(\mathbf{x})$, $i = 1,\ldots,N$. In order to close and solve the system of evolution equations (3.8) we use the approximate exponential probability density $P^*(\mathbf{x},t)$ instead of the exact solution $P(\mathbf{x},t)$ i.e.

$$P^*(\mathbf{x},t) = P_0(\mathbf{x},t)\, exp(-F^*(\mathbf{x},t)), \tag{3.13}$$

where

$$F^*(\mathbf{x},t) = \lambda_0(t) + \sum_{i=1} \lambda_i(t) f_i(\mathbf{x}), \tag{3.14}$$

where $\lambda_i(t)$, $i = 0,1,\ldots, N$ are Lagrange multipliers.

The approximate exponential probability density $P^*(\mathbf{x},t)$ is obtained by minimizing Kullback-Leibler information divergence

$$K(P(\mathbf{x},t),P_0(\mathbf{x},t)) = \int P(\mathbf{x},t)\, ln(P(\mathbf{x},t)/P_0(\mathbf{x},t)) d\mathbf{x} \geq 0 \tag{3.15}$$

under the constraints

$$\int P(\mathbf{x},t)\, d\mathbf{x} = 1, \tag{3.16}$$

and

$$\int f_i(\mathbf{x})\, P(\mathbf{x},t)\, d\mathbf{x} = \langle f_i\rangle_P, \quad i = 1,\ldots, N. \tag{3.17}$$

The method for solving this constrained optimization problem is to use the Lagrange multipliers for each of the constraints and minimize the functional

$$J = \int P\, ln(P/P_0) d\mathbf{x} + \sum_{i=0}^{N} \lambda_i(t) (\int f_i P\, d\mathbf{x} - \langle f_i\rangle_P) \tag{3.18}$$

with respect to P. Minimizing the functional (3.18) with respect to P leads to the calculation of the derivative of

$$L = P\, ln(P/P_0) + \sum_{i=0}^{N} \lambda_i(t) f_i P \tag{3.19}$$

with respect to P and setting it to the zero i.e.

$$\partial L/\partial P = ln(P/P_0) + 1 + \sum_{i=0}^{N} \lambda_i(t) f_i = 0. \tag{3.20}$$

From (3.20) one obtains the probability density P_{min} which minimizes the functional (3.18)

$$P_{min} = P_0\, exp(-(1+\lambda_0(t)) - \sum_{i=1}^{N} \lambda_i(t) f_i). \tag{3.21}$$

After replacing $1+\lambda_0(t)$ by $\lambda_0(t)$, one obtains from (3.21) the approximate exponential probability density $P^*(\mathbf{x},t)$ i.e.

$$P^* = P_0\, exp(-\lambda_0(t) - \sum_{i=1}^{N} \lambda_i(t) f_i). \tag{3.22}$$

The approximate exponential probability density $P^*(\mathbf{x},t)$ will be called the minimizing Kullback-Leibler divergence solution or short the Kullback-Leibler solution.

Inserting in (3.8) the Kullback-Leibler solution $P^*(\mathbf{x},t)$ instead of $P(\mathbf{x},t)$ we obtain

$$d\langle f_i\rangle_{P^*}/dt = \int f_i S_t (P_0 exp(-F^*))\, d\mathbf{x}, \quad i = 1,\ldots,N, \tag{3.23}$$

where

$$\langle f_i\rangle_{P^*} = \int f_i(\mathbf{x})\, P^*(\mathbf{x},t)\, d\mathbf{x}, \quad i = 1,\ldots,N. \tag{3.24}$$

Eqations (3.23) determine approximate Kullback-Leibler solutions $P^*(\mathbf{x},t)$.

In (3.23) we asume that $P(\mathbf{x},t)$ and $P^*(\mathbf{x},t)$ have the same mean values $\langle f_i\rangle_P$ and $\langle f_i\rangle_{P^*}$ for initial time. From (3.24) we may calculate $\lambda_i(t)$, $i = 1,\ldots,N$ as functions of the mean values $\langle f_i\rangle_{P^*}$, $i= 1,\ldots,N$ and $\lambda_0(t)$ is a function of $\lambda_i(t)$, $i =1,\ldots,N$ calculated from the normalization condition $\int P^*(\mathbf{x},t)\, d\mathbf{x} = 1$, then eqations (3.23) constitute a closed system of non autonomous ordinary differential equations for $\langle f_i\rangle_{P^*}$, $i= 1,\ldots,N$. Let us notice that (3.23) together with (3.13), (3.14) and (3.24) determine the differential evolution equations for $\lambda_i(t)$, $i = 1,\ldots,N$ further called the accompanied evolution equations.

Now we present formulas satisfied by the Kullback-Leibler solution $P^*(\mathbf{x},t)$ which will be useful in further considerations.

The first. From (2.2) it follows that

$$\int S_t P^*\, d\mathbf{x} = 0. \tag{3.25}$$

The second. From (3.24), (3.13) and (2.1) one gets

$$d\langle f_i\rangle_{P^*}/dt = \int f_i\, (\partial P_0/\partial t) exp(-F^*)\, d\mathbf{x} - \int f_i P_0\, exp(-F^*)(\partial F^*/\partial t)\, d\mathbf{x} = \int f_i\, (S_t P_0)\, exp(-F^*) d\mathbf{x} - \int f_i P^*(\partial F^*/\partial t)\, d\mathbf{x}, \quad i = 1,\ldots, N. \tag{3.26}$$

From (3.26) and (3.23) we finally have

$$\langle f_i \partial F^*/\partial t\rangle_{p^*} = \int f_i\, (S_t P_0)\, exp(-F^*)\, d\mathbf{x} - \int f_i S_t (P_0\, exp(-F^*))\, d\mathbf{x}, \quad i = 1,\ldots, N. \tag{3.27}$$

The third. From (3.27) it follows that

$$\langle \partial F^*/\partial t\rangle_{p^*} = \int \partial F^*/\partial t\, P^*\, d\mathbf{x} = \int (S_t P_0)\, exp(-F^*)\, d\mathbf{x}. \tag{3.28}$$

One can notice that formulas (3.25), (3.27), (3.28) which are satisfied for the Kullback-Leibler solution $P^*(\mathbf{x}, t)$ are also fulfiled for exact solution $P(\mathbf{x}, t)$ see (2.2), (3.9), (3.12).

IV. Stability of The Kullback-Leibler Solutions

We will investigate whether for the Kullback-Leibler solutions $P^*(\mathbf{x}, t)$ like for the exact solutions $P(\mathbf{x}, t)$, the generalised H-theorem (2.6) and property (2.9), i.e. $\lim_{t\to\infty}(P^*(\mathbf{x},t)-P_0(\mathbf{x},t))=0$ are satisfied.

For our consideration we take the Kulback-Leibler divergence in the following form

$$K(P^*(\mathbf{x},t),P_0(\mathbf{x},t)) = \int P^*(\mathbf{x},t)\ln(P^*(\mathbf{x},t)/P_0(\mathbf{x},t))\, d\mathbf{x} \geq 0. \quad (4.1)$$

We calculate its time derivative. According to the (3.13), (3.14), (3.28), (3.23), (3.25) one obtains

$$dK/dt = \int ((\partial P^*/\partial t)\ln(P^*/P_0) + P^*\partial \ln(P^*/P_0)/\partial t)d\mathbf{x} = -\int (\partial P^*/\partial t)\, F^*\, d\mathbf{x} - \int P^*(\partial F^*/\partial t)d\mathbf{x} =$$

$$-\int (\partial P^*/\partial t)\,(\lambda_0(t) + \sum_{i=1}^{N} \lambda_i(t)f_i(\mathbf{x}))d\mathbf{x} - \langle\partial F^*/\partial t\rangle_{p^*} =$$

$$-\sum_{i=1}^{N} \lambda_i(t)(d\langle f_i\rangle_{P^*}/dt) - \langle\partial F^*/\partial t\rangle_{p^*} =$$

$$-\sum_{i=1}^{N} \lambda_i(t)\int f_i\, S_t\,(P_0\exp(-F^*)\, d\mathbf{x} - \int (S_t\, P_0)\exp(-F^*)\, d\mathbf{x} = -\int F^*\, S_t\, P^*\, d\mathbf{x} - \int (S_t\, P_0)\exp(-F^*)\, d\mathbf{x} =$$

$$\int (S_t\, P^*)\ln(P^*/P_0)\, d\mathbf{x} - \int (S_t\, P_0)\,(P^*/P_0)\, d\mathbf{x} = \int ((S_t\, P^*)\ln(P^*/P_0) - (S_t\, P_0)\,(P^*/P_0))\, d\mathbf{x}. \quad (4.2)$$

The above formula (4.2) for approximate solutions $P^*(\mathbf{x},t)$ which do not satisfy equation (2.1) is the same as the formula (2.4) obtained exact solutions $P(\mathbf{x},t)$ of equation (2.1).

According to the (2.5) the last formula in (4.2) fulfills the following inequality

$$\int ((S_t\, P^*)\ln(P^*/P_0) - (S_t\, P_0)\,(P^*/P_0))\, d\mathbf{x} \leq 0, \quad (=0 \text{ only if } P^* = P_0). \quad (4.3)$$

Then from (4.2) and (4.3) it follows that

$$dK/dt \leq 0, \quad (=0 \text{ only if } P^* = P_0). \quad (4.4)$$

The inequality (4.4) is a generalised H-theorem for the Kullback-Leibler solutions $P^*(\mathbf{x},t)$.

Using the inequality (4.4) one can investigate an asymptotic behaviour of the Kullback-Leibler solutions $P^*(\mathbf{x},t)$. Because the Kullback-Leibler information divergence $K(P^*,P_0)$ is bounded from below and (4.4) is fulfilled, one can write

$$\lim_{t\to\infty} dK/dt = 0. \quad (4.5)$$

From (4.2), (4.3) and (4.5) it follows

$$\lim_{t\to\infty}(P^*(\mathbf{x},t)/P_0(\mathbf{x},t)) = 1. \quad (4.6)$$

Because the probability densities are normed to the unity, then according to (4.6) it follows that the difference between the two probability densities $P^* = P^*(\mathbf{x},t)$ and $P_0 = P_0(\mathbf{x},t)$ vanishes as time goes to infinity, i.e

$$\lim_{t\to\infty} (P^*(\mathbf{x},t) - P_0(\mathbf{x},t)) = 0. \quad (4.7)$$

Additionally from (4.6) and (3.13)

$$\lim_{t\to\infty} F^*(\mathbf{x},t) = 0. \quad (4.8)$$

The above considerations are done under the assumption that the Kullback-Leibler solutions $P^*(\mathbf{x},t)$ exist for any time $t > t_0$ *(t_0 is an initial time).*

Remark: It is an important result for the Kullback-Leibler divergence in this paper, that from (4.2) we have $dK/dt=\int((S_t\, P^*)\ln(P^*/P_0)-(S_t\, P_0)\,(P^*/P_0))dx$ and from (2.4) $dK/dt=\int((S_t\, P)\ln(P/P_0)-(S_t\, P_0)(P/P_0))dx$. We can see that the time derivative of Kullback- Leibler divergence for approximate solutions $P^*(x, t)$ and for exact solutions $P = P(x,t)$ are described by the same formula. According to the above, we can conclude that the derivative of the Kullback-Leibler divergence for approximate solutions and the derivative of the Kullback-Leibler divergence for exact solutons, fulfill the same inequalities (4.4) and (2.6). In consequence the approximate solutions $P^*(x, t)$ have the same asymptotic behaviour as the exact solutions $P = P(x,t)$.

V. The Accompanied Evolution Equations and The Optimization Algorithm

In order to derive the accompanied differential evolution equations for $\lambda_i(t)$, $i = 1,\ldots,N$ we start from the equations (3.27) i.e.

$$\langle f_i\partial F^*/\partial t\rangle_{p^*} = \int f_i\,(S_t\, P_0)\exp(-F^*)\, d\mathbf{x} - \int f_i\, S_t\,(P_0\exp(-F^*))\, d\mathbf{x}, \quad i=1,\ldots,N. \quad (5.1)$$

Using (3.14) on the left side of (5.1) one obtains

$$\langle f_i\,\partial F^*/\partial t\rangle_{P^*} = \langle f_i\,(d\lambda_0/dt + \sum_{j=1}^{N}(d\lambda_j/dt)f_j)\rangle_{P^*} = d\lambda_0/dt\,\langle f_i\rangle_{P^*} + \sum_{j=1}^{N}(d\lambda_j/dt)\,\langle f_i f_j\rangle_{P^*},$$

$$i=1,\ldots,N. \quad (5.2)$$

For further calculations we use formula (3.28) i.e.

$$\langle\partial F/\partial t\rangle_{p^*} = \int (S_t\, P_0)\exp(-F^*)\, d\mathbf{x}. \quad (5.3)$$

Using (3.14) on the left side of (5.3) one obtains

$$<\partial F/\partial t>_{P*} = d\lambda_0/dt + \sum_{j=1}^{N} (d\lambda_j/dt) <f_j>_{P*}. \tag{5.4}$$

From (5.3) and (5.4) we have

$$d\lambda_0/dt = \int (S_t P_0) \exp(-F^*)\, d\mathbf{x} - \sum_{j=1}^{N} (d\lambda_j/dt) <f_j>_{P*}. \tag{5.5}$$

Inserting in (5.2) instead of $d\lambda_0/dt$, the formula (5.5) we obtain

$$<f_i\, \partial F^*/\partial t>_{P*} = <f_i>_{P*} \int (S_t P_0) \exp(-F^*) d\mathbf{x} - \sum_{j=1}^{N} (d\lambda_j/dt) <f_i>_{P*} <f_j>_{P*} + \sum_{j=1}^{N} (d\lambda_j/dt) <f_i f_j>_{P*} =$$

$$\int <f_i>_{P*} (S_t P_0) \exp(-F^*) d\mathbf{x} + \sum_{j=1}^{N} M_{ij} (d\lambda_j/dt),\ i = 1, \ldots, N, \tag{5.6}$$

where

$$M_{ij} = <f_i f_j>_{P*} - <f_i>_{P*} <f_j>_{P*} \tag{5.7}$$

is a completely positive definite matrix (matrix of correlation of linearly independent functions $f_i = f_i(\mathbf{x}),\ i = 1, \ldots, N$).

Finally from (5.1) and (5.6) we obtain a system of non autonomous ordinary differential equations for $\lambda_i(t),\ i = 1, \ldots, N$ in the followig form

$$-\sum_{j=1}^{N} M_{ij} (d\lambda_j/dt) = \int (f_i\, S_t (P_0 \exp(-F^*)) - (f_i - <f_i>_{P*}) (S_t P_0) \exp(-F^*)) d\mathbf{x},\ i = 1, \ldots, N, \tag{5.8}$$

where

$$\lambda_0 = \ln\left(\int P_0 \exp\left(-\sum_{j=1}^{N} \lambda_i f_i\, d\mathbf{x}\right)\right). \tag{5.9}$$

The equation (5.9) follows from the normalisation condition $\int P^*(\mathbf{x},t)\, d\mathbf{x} = \int P_0(\mathbf{x},t) \exp(-F^*) d\mathbf{x} = 1$.

A domain $\mathbf{\Lambda}$ of the above equations is a set of such elements $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ for which all integrals in (5.8) and (5.9) exist. One can check that $\boldsymbol{\lambda} = (0, \ldots, 0) = \mathbf{0}$ is a stationary point for the system (5.8) and in this case $P^*(\mathbf{x},t) = P_0(\mathbf{x},t)$.

The system (5.8) will be called completely stable, when its every solution $\boldsymbol{\lambda}(t) = (\lambda_1(t), \ldots, \lambda_N(t))$ may be extended for any time $t > t_0$ (t_0 is an initial time) and $\lim_{t \to \infty} \boldsymbol{\lambda}(t) = \mathbf{0}$.

We may formulate two important properties of solutions of the system (5.8).

Property I. *If $\boldsymbol{\lambda}(t) = (\lambda_1(t), \ldots, \lambda_N(t))$ is a solution of the system (5.8) such that $\boldsymbol{\lambda}(t_0) \in \mathbf{\Lambda}$ (t_0 is an initial time) and $\boldsymbol{\lambda}(t)$ may be extended into domain $\mathbf{\Lambda}$ for any time $t > t_0$, then $\lim_{t \to \infty} \boldsymbol{\lambda}(t) = \mathbf{0}$, i.e. $\boldsymbol{\lambda}(t)$ tends to the stationary solution of the system (5.8).*

Let $\boldsymbol{\lambda}(t)$ be a solution of (5.8) which may be extended into domain $\mathbf{\Lambda}$ for any time $t > t_0$. For such solution according to (4.8) and (3.14) we have

$$\lim_{t \to \infty} \lambda_0(t) + \sum_{i=1}^{N} \lim_{t \to \infty} \lambda_i(t) f_i(\mathbf{x}) = 0. \tag{5.10}$$

Because functions $f_i(\mathbf{x}),\ i = 1, \ldots, N$ are linearly independent (together with $f_0(\mathbf{x}) = 1$), from (5.10) one gets $\lim_{t \to \infty} \lambda_i(t) = 0$, for $i = 0, 1, \ldots, N$. (5.11)

Property II. *In the case when $\mathbf{\Lambda} = \mathbf{E}^n$ ($\mathbf{E}^n$ is the n-dimensional Euclidean space), then the system (5.8) is completely stable.*

In the case when $\mathbf{\Lambda} \neq \mathbf{E}^n$, the system(5.8) is completely stable if every vector $d\,\boldsymbol{\lambda}/dt$, which components are given by (5.8) for every point $\boldsymbol{\lambda}$ belonging to the boundary of the set $\mathbf{\Lambda}$, is directed into $\mathbf{\Lambda}$.

The system (5.8) has the linear approximation in the following form

$$-\sum_{j=1}^{N} M_{ij}^{(0)} (d\lambda_j/dt) = \int (f_i\, S_t (P_0 \exp(-F^*)) - (f_i - <f_i>_{Po}) (S_t P_0) \exp(-F^*)) d\mathbf{x},\ i = 1, \ldots, N, \tag{5.12}$$

where

$$M_{ij}^{(0)} = <f_i f_j>_{Po} - <f_i>_{Po} <f_j>_{Po}. \tag{5.13}$$

Equations (5.12) constitute a system of linear, but in general non autonomous, ordinary differential equations.

One can notice that the function

$$V(\boldsymbol{\lambda},t) = K(P^*(\mathbf{x},t), P_0(\mathbf{x},t)) = \int P^*(\mathbf{x},t) \ln(P^*(\mathbf{x},t)/P_0(\mathbf{x},t)) d\mathbf{x} \geq 0 \tag{5.14}$$

is a Liapunov function for (5.8) [9].

Now we may formulate an approximation algorithm supported on minimizing the Kullback-Leibler information divergence in continuous systems. Approximation algorithm consists of the following steps:

Step 1. We check if a continuous system described by the system operator S_t is a Kullback system i.e. if operator S_t satisfy the inequality (2.5).

Step 2. We choose convenient set of functions $f_i = f_i(\mathbf{x}),\ i = 1, \ldots, N$ for our considerations.

Step 3. We calculate integrals on the right side of the accompanied evolution equations (5.8) using equation (5.9).

Step 4. We solve the accompanied evolution equations (5.8), (5.9) and obtain coefficients $\lambda_i(t),\ i = 0, 1, \ldots, N$.

Step 5. The coefficients $\lambda_i(t),\ i = 0, 1, \ldots, N$ are used in (3.13), which is the Kullback-Leibler solution.

VI. Conslusions

We have obtained the important result for the Kullback-Leibler divergence, that the time derivatives of the Kullback-Leibler divergence for approximate solutions $P^*(\boldsymbol{x},t)$ and for exact solutions $P = P(\boldsymbol{x},t)$ have the same shape. As a result, in the Kullback-Leibler systems the inequality $dK/dt \leq 0$ is satisfied for exact and approximate solutions. In consequence, the approximate solutions $P^*(\boldsymbol{x},t)$ have the same asymptotic behaviour as the exact solutions $P = P(\boldsymbol{x},t)$ i.e. they tend to the same solution P_0.
We have created an approximation algorithm supported on minimizing the Kullback-Leibler information divergence in the Kullback-Leibler systems. Practical application of the proposed approach requires knowledge of the probability density P_0. If a Kullback system possesses a stationary solution, this solution may be chosen as P_0 [4]. In the case when the Kullback system possesses a time-dependent periodic solution [6], this periodic solution is an attractor and may be chosen as P_0. Let us notice that the approximation algorithm presented in this paper not only gives a certain approximate scheme of solving the Kullback system but also generalizes the information gain minimizing approach for the Fokker-Planck equation, even with time dependent drift and diffusion coefficients[6].
Problems analogical to the ones presented in this paper were investigated by the author in the case of discrete systems and will be published in a separate paper.

Appendix

Here we will show that the system described by the Fokker-Planck Equation (F.P.E.) is a Kullback system.
We take for our consideration the n-dimensional F.P.E. with time-dependent drift and diffusion coefficients for the probability density function $P(\boldsymbol{x},t)$ of the continuous random variable $\boldsymbol{x} = (x_1,\ldots, x_n)$ in the following form[8]

$$\partial P/\partial t = -\sum_{i=1}^{n} \partial(v_i P)/\partial x_i + \sum_{i,j=1}^{n} \partial(D_{ij}(\partial P/\partial x_j))/\partial x_i, \qquad \text{(A.1)}$$

were $v_i = v_i(\boldsymbol{x},t)$ is a drift vector and $D_{ij} = D_{ij}(\boldsymbol{x},t)$ is a symmetric and completely positive definite diffusion matrix[1,7]. The system operator S_t in the case of F.P.E. will be denoted as S_t^{FPE} and is defined below

$$S_t^{FPE} P = -\sum_{i=1}^{n} \partial(v_i P)/\partial x_i + \sum_{i,j=1}^{n} \partial(D_{ij}(\partial P/\partial x_j))/\partial x_i . \qquad \text{(A.2)}$$

We will check that the formula (2.4) is satisfied for the system operator S_t^{FPE} i.e.

$$\int((S_t^{FPE} P)ln(P/P_0) - (S_t^{FPE} P_0)(P/P_0))d\boldsymbol{x} \leq 0 \text{ , (=0 only if } P = P_0). \qquad \text{(A.3)}$$

Substituting (A.2) on the left side of (A.3) one obtains

$$\int(ln(P/P_0)(-\sum_{i=1}^{n} \partial(v_i P)/\partial x_i + \sum_{i,j=1}^{n}\partial(D_{ij}(\partial P/\partial x_j))/\partial x_i) - (P/P_0)(-\sum \partial(v_i P_0)/\partial x_i + \sum \partial(D_{ij}(\partial P_0/\partial x_j))/\partial x_i))d\boldsymbol{x}$$

$$= -\int \sum_{i=1}^{n} ln(P/P_0)\partial(v_i P)/\partial x_i d\boldsymbol{x} + \int \sum_{i,j=1}^{n} ln(P/P_0)\partial(D_{ij}(\partial P/\partial x_j))/\partial x_i d\boldsymbol{x} + \int \sum_{i=1}^{n}(P/P_0)\, \partial(v_i P_0)/\partial x_i d\boldsymbol{x}$$

$$-\int \sum_{i,j=1}^{n} (P/P_0)\partial(D_{ij}(\partial P_0/\partial x_j))/\partial x_i d\boldsymbol{x} = \int \sum_{i=1}^{n}\partial(ln(P/P_0))/\partial x_i(v_i P)d\boldsymbol{x} - \int \sum_{i,j=1}^{n}\partial(ln(P/P_0))/\partial x_i (D_{ij}(\partial P/\partial x_j))d\boldsymbol{x}$$

$$-\int \sum_{i=1}^{n}\partial(P/P_0)/\partial x_i (v_i P_0)d\boldsymbol{x} + \int \sum_{i,j=1}^{n}\partial(P/P_0)/\partial x_i(D_{ij}(\partial P_0/\partial x_j))d\boldsymbol{x} = \int \sum_{i=1}^{n}(P_0/P)\partial(P/P_0)/\partial x_i(v_i P)d\boldsymbol{x}$$

$$-\int \sum_{i,j=1}^{n}\partial(ln(P/P_0))/\partial x_i (D_{ij}(\partial((P/P_0)P_0)/\partial x_j))d\boldsymbol{x} - \int \sum_{i=1}^{n}\partial(P/P_0)/\partial x_i (v_i P_0)d\boldsymbol{x}$$

$$+\int \sum_{i,j=1}^{n}\partial(P/P_0)/\partial x_i(D_{ij}(\partial P_0/\partial x_j))d\boldsymbol{x} = -\int \sum_{i,j=1}^{n}\partial(ln(P/P_0))/\partial x_i (D_{ij}(\partial(P/P_0)/\partial x_j P_0 + (P/P_0)\partial P_0/\partial x_j)d\boldsymbol{x}$$

$$+\int \sum_{i,j=1}^{n}\partial(P/P_0)/\partial x_i(D_{ij}(\partial P_0/\partial x_j))d\boldsymbol{x} = -\int \sum_{i,j=1}^{n} (P_0/P)\partial(P/P_0)/\partial x_i (D_{ij}(\partial(P/P_0)/\partial x_j P_0 + (P/P_0)\partial P_0/\partial x_j)d\boldsymbol{x}$$

$$+\int \sum_{i,j=1}^{n}\partial(P/P_0)/\partial x_i(D_{ij}(\partial P_0/\partial x_j))d\boldsymbol{x} = -\int \sum_{i,j=1}^{n} (P_0/P)\partial(P/P_0)/\partial x_i D_{ij}(\partial(P/P_0)/\partial x_j(P_0/P)\, P\, d\boldsymbol{x}$$

$$-\int^{n} \sum_{i,j=1}\partial(P/P_0)/\partial x_i D_{ij}\, \partial P_0/\partial x_j)d\boldsymbol{x} + \int^{n} \sum_{i,j=1}\partial(P/P_0)/\partial x_i(D_{ij}(\partial P_0/\partial x_j))d\boldsymbol{x} =$$

$$-\int \sum_{i,j=1}^{n} \partial ln(P/P_0)/\partial x_i D_{ij}\partial ln(P/P_0)/\partial x_j P\, d\boldsymbol{x}. \qquad \text{(A.4)}$$

Let us notice, in connection with the above calculations in (A.4), that adjoint manipulation associated with spatial operations on probability density function $P(\boldsymbol{x},t)$ is possible only if $P(\boldsymbol{x},t)$ is rapidly decreasing for $|\boldsymbol{x}| \to \infty$ (the natural boundary condition according to Graham [1]).
Because D_{ij} is a completely positive definite matrix, the last formula in (A.4) satisfies the following inequality

$$-\int \sum_{i,j=1}^{n} \partial ln(P/P_0)/\partial x_i D_{ij}\partial ln(P/P_0)/\partial x_j P\, d\boldsymbol{x} \leq 0 \text{ , (=0}$$

only if $P = P_0$).

(A.5)

From (A.5), (A.4) it follows that inequality (A.3) is fulfilled for the system operator S_t^{FPE}, so it is the Kullback system operator. Hence our earlier general considerations connected to the Kullback systems may be applied for the systems described by F.P.E..

References

[1] R. Graham, “Springer Tract in Modern Physics”, Springer-Verlag, Berlin, vol. 66, 1973.
[2] S. Kullback, R.Leibler, “On information and sufficiency”, Ann. Math. Stat., vol. 22, pp 79-86, 1951.
[3] S. Kullback, “Information theory and statistic”, Wiley, New York,1959.
[4] J. Owedyk, “ Investigation of stability of information gain solutions of a Fokker-Planck equation”, Acta Phys. Polon., Vol. A63, no. 3, pp (317-327), 1983.
[5] J. Owedyk, “Stability of a stationary distribution of a Master Equation with respect to information gain solutions”, Z. Phys. B, vol. 54, pp (183-186), 1984.
[6] J. Owedyk, “On the Fokker-Plank equation with time-dependent drift diffusion coefficients and its exponential solutions”, Z. Phys. B, vol. 59, pp (69-74),1985.
[7] J. Owedyk, “On the master equation with time-dependent transition probabilities”, Phys. Lett., vol 109, pp (152-154), issue 20 may 1985.
[8] H. Risken, “The Fokker-Planck equation”, Springer-Verlag, Berlin, 1996.
[9] J. la Salle ,S. Lefschetz, “Stability by Liapunov’s direct method”, New York, London, Academic Press,1961.