Journal of Applied Data Analysis and Modern Stochastic Modelling
Vol. 1 No. 2 (December, 2024)

Received: 20th June 2024 *Revised: 25th August 2024* *Accepted: 17th October 2024*

STOCHASTIC MODELS FOR SYSTEMS OF FORWARD KOLMOGOROV EQUATIONS

YANA BELOPOLSKAYA*

Abstract. Stochastic counterparts of the Cauchy problem for systems of nonlinear second order parabolic equations written in terms of forward SDEs and FBSDEs are derived and studied. There are selected two types of nonlinear PDE systems arising in applications and there are developed two different stochastic approaches to study them. The stochastic approach to systems of the first type allows to develop a numerical algorithm to construct the Cauchy problem solutions based on the suitable FBSDE and the neural network technique. The stochastic approach to systems of the second type extends the results obtained in the theory of McKean-Vlasov type equations.

1. Introduction

In this paper we derive stochastic interpretations of the Cauchy problem for several classes of systems of nonlinear parabolic equations.

Namely, we consider the forward Cauchy problem for systems of nonlinear parabolic equations and show that one can construct probabilistic representations of their solutions. We note that probabilistic interpretation could be applied both to better understanding of a "physical" nature of the PDE system under consideration and as a mere effective tool to construct an effective numerical scheme to obtain an approximation of a solution to the PDE system. We put the word physical in quotes since it is true as well for systems describing chemical, biological or financial objects.

In this paper we review some known stochastic models associated with the forward Cauchy problem for systems of nonlinear parabolic equations and suggest a new one. All the systems considered here belong to the class of the so called reaction-diffusion systems arising as mathematical models in various fields, namely, in physics, chemistry, biology and so on and were studied by many authors (see [1], [2] and references there).

As a matter of fact we consider stochastic systems which underline the forward Cauchy problem solutions for two types of nonlinear PDE systems. A system of the first type can be easily reduced to the backward Cauchy problem for a suitable PDE system while a system of the second type does not allow such a reduction.

Date: Date of Submission February 01, 2024 ; Date of Acceptance February 05, 2024, Communicated by Mamadsho Ilolov.

2000 *Mathematics Subject Classification.* Primary 60H10; Secondary 69H35, 35K57.

Key words and phrases. nonlinear forward Kolmogorov equations and systems, stochastic models, numerical algorithms, neural networks.

The first type of second order parabolic equation systems includes systems of the form

$$\frac{\partial u_m}{\partial t} = \frac{1}{2}TrB(x,u)\nabla^2 u_m + a_m(x,u)\cdot\nabla u_m + \sum_{q=1}^{d_1}\sum_{i=1}^{d} C^i_{mq}(x,u)\nabla_i u_q + \tag{1.1}$$

$$+\sum_{q=1}^{d_1} c_{mq}(u)u_q = f(t,x,u), \quad u_m(0,x) = u_{0m}(x), x\in R^d,\ m = 1,\dots,d_1,$$

and

$$\frac{\partial v_m}{\partial t} = \frac{1}{2}TrB_m(x,v)\nabla^2 v_m + a_m(x,v)\cdot\nabla v_m + \sum_{q=1}^{d_1} c_{mq}(x,v)v_q = 0, \tag{1.2}$$

$$v_m(0,x) = v_{0m}(x).$$

Here d, d_1 are integers, $u(t,x)\in R^{d_1}$, $x\in R^d$ $t\in[0,T]$,

$$\nabla u(x) = \{\nabla_i u(x)\}_{i=1}^d, \quad \nabla^2 u = \{\nabla^2_{y_iy_j}u\}_{i,j=1}^d, \quad TrB\nabla^2 u = \sum_{i,j=1}^d B_{ij}\nabla^2_{y_iy_j}u,$$

$x\cdot y = \sum_{i=1}^d x_i y_i$ is the inner product in R^d, $B = AA^*\geq 0$, where A^* denotes a dual matrix.

The forward Cauchy problem (1.1), (1.2) could be easily reduced to the backward Cauchy problem for suitable systems of parabolic equations

$$\frac{\partial g_m}{\partial t} + \frac{1}{2}TrB(x,g)\nabla^2 g_m + a_m(x,g)\cdot\nabla g_m + \sum_{q=1}^{d_1}\sum_{i=1}^{d} C^i_{mq}(x,g)\nabla_i g_q + \tag{1.3}$$

$$+\sum_{q=1}^{d_1} c_{mq}(x,g)g_q = f_m(t,x,g), \quad g_m(T,x) = u_{0m}(x),$$

and

$$\frac{\partial f_m}{\partial t} + \frac{1}{2}TrB_m(x,f)\nabla^2 f_m + a_m(x,f)\cdot\nabla u_m + \sum_{q=1}^{d_1} c_{mq}(x,f)f_q = 0, \tag{1.4}$$

$$f_m(T,x) = u_{0m}(x)$$

respectively. To this end it is enough to set

$$u_m(t,x) = g_m(T-t,x), \quad v_m(t,x) = f_m(T-t,x).$$

Stochastic approach to investigation of scalar equations of this type called semilinear parabolic equations was developed by M.Freidlin [3]. This approach was extended to systems of semilinear parabolic equations by by Yu.Dalecky and Ya.Belopolskaya [4], [5]. In addition, it was extended in [6], to systems of quasilinear and fully nonlinear parabolic equations and systems. We say that an equation

$$\frac{\partial v}{\partial t} = \frac{1}{2}TrB^v(x)\nabla^2 v + a^v(x)\cdot\nabla v + c^v(x)v, \quad v(0,x) = u_0(x)$$

is semilinear, quasilinear or fully nonlinear if its coefficients have the form

$$a^v(x) = a(x,v), \quad a^v(x) = a(x,v,\nabla v) \quad \text{or} \quad a^v(x) = a(x,v,\nabla v,\nabla^2 v)$$

respectively.

An alternative approach to quasilinear equations was suggested by E.Pardoux and S.Peng [7], [8] and developed by many authors (see [9] –[11] and references there). This approach was applied to systems of the form (1.3) in [12] and to systems of the form (1.4) in [8].

The second type of systems includes systems of the form

$$\frac{\partial \mu_m}{\partial t} = \frac{1}{2}\sum_{q=1}^{d_1} \Delta\left(B_{mq}[y,\mu_q]\mu_m\right) - \sum_{q=1}^{d_1} div(b_{mq}[y,\mu_q]\mu_m)$$

$$+\sum_{q=1}^{d_1} c_{mq}[y,\mu_q]\mu_m, \quad \mu_m(0,dy) = u_{0m}(y)dy, \tag{1.5}$$

and

$$\frac{\partial \mu_m}{\partial t} = \frac{1}{2} div\left[\sum_{q=1}^{d_1} B_{mq}[y,\mu_q]\nabla\mu_m\right] + \sum_{q=1}^{d_1} c_{mq}(u)\mu_q \tag{1.6}$$

$$\mu_m(0,y) = u_0(y)dy, y \in R^d.$$

Here $B^{ij}_{mq}(y,\mu)$ is the forth order tensor $m = 1,\dots,d_1, i,j = 1,\dots,d$ and we use notation of the type

$$B_{mq}[y,\mu_q] = \int_{R^d} B_{mq}(y-x)\mu_q(dy).$$

We show that one can construct stochastic models associated with systems of both types and derive probabilistic representations of the Cauchy problem solutions for these systems.

The pioneer paper concerning connections between the forward Cauchy problem (1.5) for the case $d_1 = 1$ is due to H.McKean [13]. The theory was extended in a number of papers (see references in [14].) In these papers there were constructed stochastic processes which allowed to derive probabilistic representations for the required solutions of the Cauchy problem for semilinear parabolic equations under consideration.

Let us briefly recall the both approaches.

The Freidlin's approach to the Cauchy problem for a forward semilinear parabolic equation

$$\frac{\partial v}{\partial t} = \frac{1}{2} TrA(x,v)\nabla^2 vA^*(x,v) + a(x,v)\cdot\nabla v + c(x,v)v, \quad v(0,x) = u_0(x) \tag{1.7}$$

is based on the possibility to reduce (1.7) to the backward Cauchy problem

$$\frac{\partial u}{\partial t} + \frac{1}{2} TrA(x,u)\nabla^2 uA^*(x,u) + a(x,u)\cdot\nabla u + c(x,u)u = 0, \quad u(T,x) = u_0(x). \tag{1.8}$$

setting $u(T-t,x) = v(t,x)$.

To construct a stochastic model associated with (1.8) we fix a probability space $(\Omega,\mathcal{F},P)$ and denote by $w(t)$ the Wiener process valued in R^d defined on this probability space. Let $\mathcal{F}_t \subset \mathcal{F}$ be a flow of σ- sub-algebras of $\mathcal{F}$ generated by $w(t)$.

Given a fixed time $T > 0$ one can interpret (1.8) as the backward Kolmogorov equation for a stochastic process satisfying the stochastic differential equation (SDE)

$$d\xi(\theta) = a(\xi(\theta), u(T-\theta, \xi(\theta)))d\theta + A(\xi(\theta), u(T-\theta, \xi(\theta)))dw(\theta), \quad \xi(t) = x, \tag{1.9}$$

where

$$u(T-t, x) = E\left[u_0(\xi_{t,x}(T))e^{\int_t^T c(\xi_{t,x}(\tau), u(T-\tau, \xi_{t,x}(\tau)))d\tau}\right]. \tag{1.10}$$

The stochastic representation (1.10) of the solution $u(T-t,x)$ to the Cauchy problem (1.8) allows to reduce the original Cauchy problem (1.8) to an integral equation (1.10). Next one has to state conditions on coefficients in (1.9) and the function u_0 which allow to find a suitable functional space where a nonlinear map $\Phi_T(t,x,u) = u(T-t,x)$ generated by this integral equation is a contraction. If moreover one can prove that in addition the function $u(t,y)$ satisfying (1.10) is twice differentiable then he can state that it is a unique classical solution to (1.8).

The investigation of nonlinear forward Kolmogorov equations and systems of the form (1.5), (1.6) is a more delicate thing. First, due to the probabilistic interpretation to be described below, we know that solutions of these equations should be either measures or measure densities. If we look for a solution of the system (1.3) in a space of measures then the coefficients of the equations should be nonlinear functionals of the solution.

The approach developed by McKean [13] allows to construct a stochastic model associated with the Cauchy problem for the Vlasov equation

$$\frac{\partial u(t,y)}{\partial t} = \frac{1}{2}Tr\nabla^2(B[y,u]u(t,y)) - div(a[y,u]u(t,y)), \quad u(0,y) = u_0(y). \tag{1.11}$$

arising in plasma physics. Here $Tr\nabla^2(B[y,u]u) = \sum_{i=1}^d \frac{\partial^2(B_{ij}[y,u]u)}{\partial y_i \partial y_j}$, $B_{ij}[y.u] = \sum_{k=1}^d A_{ik}[y,u]A_{kj}[y,u]$ is a positive matrix, $div\, a[y,u] = \sum_{i=1}^d \nabla_{y_i} a_i[y,u]$ and coefficients $A[y,u] \in R^d \otimes R^d, a[y,u] \in R^d$ have the form

$$a_i[y,u] = \int_{R^d} \cdots \int_{R^d} a_i(y, y_1, \ldots, y_d)u(y_1)\ldots u(y_d)dy_1 \ldots dy_d, \quad i = 1, \ldots, d. \tag{1.12}$$

The parabolic equation (1.11) can be treated as a forward Kolmogorov equation for a distribution $u(t,dy) = P\{\xi(t) \in dy\}$ of a stochastic process $\xi(t) \in R^d$ governed by a stochastic equation

$$\xi(t) = \xi(0) + \int_0^t \left[\int_R \cdots \int_R a(\xi(s), y_1, \ldots, y_d)u(s,dy_1)\ldots u(s,dy_d)\right] ds +$$

$$+ \int_0^t \left[\int_{R^d} \cdots \int_{R^d} A(\xi(s), y_1, \ldots, y_d)u(s,dy_1)\ldots u(s,dy_d)\right] dw(s), \tag{1.13}$$

where $\xi(0) \in R^d$ does not depend on $w(t)$ and $P(\xi(0) \in dy) = u_0(y)dy$.

Two stochastic approaches to the forward Cauchy problem for systems of nonlinear parabolic equations (1.3) were developed in [5], [12]. One of them is based on forward SDEs associated with the system and the other is based on forward-backward SDEs (FBSDEs) associated with it. Both approaches can be useful to develop effective algorithms to construct numerical approximations of classical and viscosity solutions of the Cauchy problem for systems of nonlinear PDEs [15], [16].

In this paper we describe an approach to construct an algorithm to obtain a numerical solution of (1.1) based on the FBSDE underlying (1.3). It extends the approaches suggested in recent papers [17]– [20].

We consider as well a stochastic interpretation of systems (1.5) and (1.6).

The remaining part of the paper is structured as follows. In section 2 we develop and study stochastic counterparts of the Cauchy problem (1.3) in terms of forward SDEs and FBSDEs.

In section 3 we develop a numerical algorithm to construct the Cauchy problem solution for (1.3) based on its connection with the FBSDE and using neural network technique.

In section 4 we derive a stochastic representation of the systems (1.5) and (1.6).

Finally in section 5 we study the derived stochastic counterparts of (1.5) and (1.6) and their connection with the Cauchy problem for PDEs.

2. Stochastic models of the forward Cauchy problem for type 1 systems

Denote by $C_0^\infty(R^d)$ ($C_0^\infty([0,T]\times R^d)$) the space of infinite differentiable functions defined on R^d (on $[0,T]\times R^d$) with compact supports and let $C_b(R^d)$ be the space of bounded continuous functions $f : R^d \to R$ with the norm $\|f\|_\infty = \sup_x |f(x)|$. We use notations $C^k(R^d)$ for the space of k-times differentiable real valued functions defined on R^d.

To simplify notations below we assume the Einstein convention about summing over the repeating indices if the contrary is not mentioned.

Consider the forward Cauchy problem for systems of the form

$$\frac{\partial u_m}{\partial t} = \frac{1}{2} TrB(y,u)\nabla^2 u_m + a(y,u)\cdot\nabla u_m + F^k_{mq}(y,u)\nabla_{y_k} u_q + c_{mq}(y,u)u_q, \tag{2.1}$$

$$u_m(0,y) = u_{0m}(y), \quad m,q = 1,\dots,d_1,$$

where $B = AA^* > 0$ is a positive definite matrix, $F^k_{mq} = \sum_{j=1}^d C^j_{mq}A^{jk}$ and

$$\frac{\partial g_m}{\partial t} = \frac{1}{2} TrB_m(y,g)\nabla^2 g_m + a_m(y,g)\cdot\nabla g_m + c_{mq}(y,g)g_q, \tag{2.2}$$

$$g_m(0,y) = g_{0m}(y).$$

Notice that in (2.2) the summation in m is not assumed.

Similar to the scalar case described in the introduction one can easily verify that functions $v_m(T-t,y) = u_m(t,y)$ satisfy the backward Cauchy problem

$$\frac{\partial v_m}{\partial t} + \frac{1}{2} TrB(y,v)\nabla^2 v_m + a(y,v)\cdot\nabla v_m + F^k_{mq}(y,v)\nabla_{y_k} v_q + c_{mq}(y,v)v_q = 0, \tag{2.3}$$

$$v_m(T,y) = u_{0m}(y), \quad m,q = 1,\dots,d_1$$

and a similar transformation $g_m(t,y) = f_m(T-t,y)$ allows to reduce the Cauchy problem (2.2) to the corresponding backward Cauchy problem.

$$\frac{\partial f_m}{\partial t} + \frac{1}{2} TrB_m(y,f)\nabla^2 f_m + a_m(y,f)\cdot\nabla f_m + c_{mq}(y,f)f_q = 0, \tag{2.4}$$

$$f_m(T,y) = g_{0m}(y), \quad m = 1,\dots,d_1.$$

Note that both of these systems (2.3) and (2.4) admit a reduction to scalar equations. Namely, the system (2.3) can be treated as a scalar equation with respect to the function $\Phi(t,y,h) = h\cdot v(T-t,y)$ defined on $[0,T]\times R^d\times R^{d_1}$, $h\in R^{d_1}$. The system (2.4) can be treated as a scalar equation with respect to a function $f(T-t,y,m) = f_m(T-t,y)$, $m\in V=\{1,2,\dots,d_1\}$ defined on $[0,T]\times R^d\times V$. This comes to be obvious when one looks on probabilistic representations of solutions to (2.3) and (2.4) (see [5],[8], [**?**]).

To obtain a probabilistic representation of a classical solution of the Cauchy problem (2.3) we consider a stochastic problem

$$d\xi(\theta) = a(\xi(\theta), v(T-\theta,\xi(\theta)))d\theta + A(\xi(\theta), v(T-\theta,\xi(\theta)))dw(\theta), \quad \xi(t)=y, \tag{2.5}$$

$$d\eta(\theta) = c^*(\xi(\theta), v(T-\theta,\xi(\theta)))\eta(\theta)d\theta + C^*(\xi(\theta), v(T-\theta,\xi(\theta)))(\eta(\theta), dw(\theta)), \tag{2.6}$$

$$h\cdot v(T-t,y) = E[\eta_{t,h}(T)\cdot u_0(\xi_{t,y}(T))], \qquad \eta(t)=h. \tag{2.7}$$

The existence and uniqueness of a solution to the system (2.5) – (2.7) were proved in [5]. Besides it was proved that if coefficients and initial data are smooth enough then (2.7) defines a unique classical solution of the Cauchy problem (2.3) .

The representation (2.7) prompts that the system (2.3) can be considered as a scalar equation w.r.t. $\Phi(T-t,\gamma) = h\cdot v(T-t,y)\rangle, \quad \gamma=(y,h)$

$$\frac{\partial\Phi}{\partial t} + \frac12 TrQ(\gamma,\Phi)\nabla^2_\gamma\Phi Q^*(\gamma,\Phi) + q(\gamma,\Phi)\cdot\nabla_\gamma\Phi = 0 \tag{2.8}$$

and

$$\Phi(T,\gamma) = \Phi_0(\gamma) = h\cdot u_0(\xi(T)), \quad \gamma=(y,h).$$

Here $q(\gamma)$ and $Q(\gamma)$ are given by

$$q(\gamma,\Phi) = \begin{pmatrix} a(x,v) \\ c(x,v)h \end{pmatrix},$$

$$Q(\gamma,\Phi)Q^*(\gamma,\Phi) = \begin{pmatrix} A(x,v)A^*(x,v) & A(x,v)C(x,v)h \\ A(x,v)C(x,v)h & M \end{pmatrix},$$

where M is arbitrary bounded function such that QQ^* is a positive matrix. One can check that

$$\frac12 Tr\,Q(\gamma,\Phi)\nabla^2\Phi Q^*(\gamma,\Phi) =$$
$$h\cdot\frac12 Tr\begin{pmatrix} A(x,v)A^*(x,v) & C(x,v)A(x,v) \\ C(x,v)A(x,v) & M \end{pmatrix}\begin{pmatrix} \nabla^2 v & \nabla v \\ \nabla v & 0 \end{pmatrix} =$$
$$\frac12 h\cdot TrA(x,v)A^*(x,v)\nabla^2 v(x) + C^*(x,v)h\cdot A(x,v)\nabla v,$$
$$q(\gamma,\Phi)\cdot\nabla_\gamma\Phi(\gamma) = h\cdot a(x,v)\nabla v + c^*(x,v)h\cdot v(x).$$

Let us briefly show that this point of view allows to apply the FBSDE theory to construct a viscosity solution of a quasilinear system of the form (2.3) (see details in [12]).

Given $\Phi(t,\gamma) = h\cdot u(t,y)$ we denote

$$\|\Phi(t,\gamma)\|_1 = \sup_{\{h:\|h\|=1, y\in R^d\}} |h\cdot u| = \|u\|.$$

Consider stochastic processes

$$\gamma(t) = (\xi(t), \gamma(t)), \quad Y(t) = \Phi(T-t, \gamma(t)) = \eta(t) \cdot v(T-t, \xi(t))$$

and

$$Z(t) = \nabla \Phi(t, \gamma(t)),$$

where $\xi(t)$ and $\eta(t)$ satisfy stochastic equations of the form (2.5), (2.6) with coefficients depending on $(x, v, \nabla v)$. Then we obtain an equation of the form

$$d\gamma(t) = q(\gamma, Y(t), Z(t))dt + Q(\gamma, Y(t), Z(t))dW(t), \gamma(s) = (x, h) \tag{2.9}$$

where $W(t) = (w(t), w(t))^* \in R^{2d}$. Assume that $\Phi(t, \gamma)$ is twice differentiable in γ and apply the Ito formula to obtain $d\Phi(t, \gamma(t))$. As a result we get a relation

$$\Phi(T, \gamma(T)) - \Phi(t, \gamma(t)) = \int_t^T [\frac{\partial \Phi(\tau, \gamma(\tau))}{\partial \tau} \tag{2.10}$$

$$+\frac{1}{2} Tr\, Q(\gamma(\tau), \Phi(\tau, \gamma(\tau)), \nabla \Phi(\tau, \gamma(\tau))) \nabla^2 \Phi(\tau, \gamma(\tau))) Q^*(\gamma(\tau), \Phi(\tau, \gamma(\tau)),$$
$$\nabla \Phi(\tau, \gamma(\tau))) + q(\gamma(\tau), \Phi(\tau, \gamma(\tau)), \nabla \Phi(\tau, \gamma(\tau))) \cdot \nabla \Phi(\tau, \gamma(\tau))] d\tau +$$
$$\int_t^T Q(\gamma(\tau), \Phi(\tau, \gamma(\tau)), \nabla \Phi(\tau, \gamma(\tau))) \nabla \Phi(\tau, \gamma(\tau)) \cdot dW(\tau)).$$

Since $Y(\tau) = \Phi(\tau, \gamma(\tau))$ and $Z(\tau) = \nabla \Phi(\tau, \gamma(\tau))$ we deduce from (2.8) and (2.10) that

$$Y(t) = Y(T) - \int_t^T Q(\gamma(\tau), Y(\tau), Z(\tau)) Z(\tau) \cdot dW(\tau), \tag{2.11}$$

where $Y(T) = \Phi_0(\gamma(T)) = \eta(T) \cdot v_0(\xi(T))$. As a result we get a BSDE

$$dY(t) = Q(\gamma(t), Y(t), Z(t)) Z(t) \cdot dW(t), \quad Y(T) = \eta(T) \cdot v_0(\xi(T)) \tag{2.12}$$

associated with (2.8).

One can extend this approach to systems of quasilinear parabolic equations of the form

$$\frac{\partial u_m}{\partial t} + \frac{1}{2} Tr B(y, u, \nabla u) \nabla^2 u_m + a(y, u, \nabla u) \cdot \nabla u_m + F^k_{mq}(y, u, \nabla u) \nabla_{y_k} u_q + \tag{2.13}$$
$$c_{mq}(y, u, \nabla u) u_q + L_m(y, u, \nabla u) = 0,$$
$$u_m(T, y) = u_{0m}(y), \quad m, q = 1, \ldots, d_1.$$

Repeating the previous arguments we reduce the Cauchy problem (2.13) to the scalar Cauchy problem

$$\frac{\partial \Phi}{\partial t} + \frac{1}{2} Tr Q(\gamma, \Phi, \nabla \Phi) \nabla^2_\gamma \Phi Q^*(\gamma, \Phi, \nabla \Phi) + q(\gamma, \Phi, \nabla \Phi) \cdot \nabla_\gamma \Phi + M(\gamma, \Phi, \nabla \Phi) = 0 \tag{2.14}$$

and

$$\Phi(T, \gamma) = \Phi_0(\gamma) = h \cdot u_0(\xi(T)), \quad \gamma = (y, h).$$

Here $q(\gamma, \Phi, \nabla \Phi)$ and $Q(\gamma, \Phi, \nabla \Phi)$ are given by

$$q(\gamma, \Phi, \nabla \Phi) = \begin{pmatrix} a(x, v, \nabla v) \\ c(x, v, \nabla v) h \end{pmatrix}, \quad Q(\gamma, \Phi, \nabla \Phi) Q^*(\gamma, \Phi, \nabla \Phi) =$$
$$= \begin{pmatrix} A(x, v, \nabla v) A^*(x, v, \nabla v) & A(x, v, \nabla v) C(x, v, \nabla v) h \\ A(x, v, \nabla v) C(x, v, \nabla v) h & 0 \end{pmatrix}.$$

Finally, we consider the system (2.14) and describe its stochastic counterpart

$$d\gamma(\tau) = q(\gamma(\tau), Y(\tau), Z(\tau))d\tau + Q(\gamma(\tau), Y(\tau), Z(\tau))dW(\tau), \gamma(t) = (x, h), \quad (2.15)$$

$$dY(\tau) = -M(\gamma(\tau), Y(\tau), Z(\tau))d\tau + Q(\gamma(\tau), Y(\tau), Z(\tau))Z(\tau) \cdot dW(\tau), \quad (2.16)$$
$$Y(T) = \eta(T) \cdot v_0(\xi(T)),$$

where $M(\gamma, Y, Z) = h \cdot L(y, Y, Z)$.

To obtain a closed system which allow to define three unknown processes $\xi(t), Y(t)$ and $Z(t)$ satisfying (2.15), (2.16) we consider the square integrable martingale

$$N(t) = E[Y(T) + \int_0^T M(\gamma(\tau), Y(\tau), Z(\tau))d\tau|\mathcal{F}_t].$$

Then we apply the Ito theorem which states the existence of a unique process $Z(t)$ such that

$$N(T) = E[N(T)] + \int_0^T Z(\tau)dW(\tau). \quad (2.17)$$

As a result we get that (2.15)-(2.17) is a closed system. To define a solution to this system we need some additional notations.

Denote by $n = d + d_1, \theta = 2d(d+1)$ and by:

$S^2_T(R^n)$ the set of all $\mathcal{F}_t$ adapted R^n valued processes with the norm

$$\|\gamma\|^T_S = E[\sup_{t\in[0,T]} \|\gamma(t)\|^2] < \infty,$$

$H^2_T(R^\theta)$ the set of R^θ-valued $\mathcal{F}_t$ adapted processes with the norm

$$\|Z\|^T_H = \left(E\left[\int_0^T \|Z(t)\|^2 dt\right]\right)^{\frac{1}{2}} < \infty;$$

$L^2_t(\Omega, P)$ the set of $\mathcal{F}_t-$measurable random variables such that $E\|\gamma(t)\|^2 < \infty$.

We consider a set $\mathcal{Q}$ of functions of the form $\Phi(y, h) = h \cdot u(y)$ defined on $R^d \times R^{d_1}$ such that for $\gamma = (y, h) \in R^n$,

$$\|\Phi\|_Q = \sup_{h:\|h\|=1} \sup_{y\in R^d} |h \cdot u(y)| = \sup_{y\in R^d} \|u(y)\| < \infty.$$

Further we define by $\mathcal{Z}^T = L^2(\Omega : C([0,T]; R^n) \times L^2(\Omega; C([0,T]; R)) \times H^2_T(R^\theta))$ the Banach space with the norm

$$\|(\gamma(\cdot), Y(\cdot), Z(\cdot))\|_{\mathcal{Z}} = \left\{\|\gamma\|^T_S + \|Y\|^T_S + \|Z\|^T_H\right\}^{\frac{1}{2}}$$

for all $(\gamma(\cdot), Y(\cdot), Z(\cdot)) \in \mathcal{Z}^T$.

A process $(\gamma(\cdot), Y(\cdot), Z(\cdot)) \in \mathcal{Z}^T$ is called a solution of (2.15), (2.16) if

$$\gamma(t) = \gamma + \int_0^t q(\gamma(\tau), Y(\tau), Z(\tau))d\tau + \int_0^t Q(\gamma(\tau), Y(\tau), Z(\tau))dW(\tau), \gamma(0) = (y, h) \quad (2.18)$$

$$Y(t) = \Phi_0(\gamma(T)) + \int_t^T M(\gamma(\tau), Y(\tau), Z(\tau))d\tau \quad (2.19)$$
$$+ \int_t^T Q(\gamma(\tau), Y(\tau), Z(\tau))Z(\tau) \cdot dW(\tau),$$

almost surely.

Denote by $m = (\gamma, Y, Z) \in R^n \times R \times R^\theta, \Gamma = (q, Q, L)^*$. We say that condition **C2.1** holds if for $m = (\gamma, Y, Z)$:

1. $q(m)$ $Q(m)$ and $M(m)$ have a sublinear growth;
2. $\forall \kappa \in R^d \otimes R^{d_1} \times R \times R^{2d}, \Gamma(\kappa)$ is uniformly Lipschitz continuous;
3. $u_0(x)$ is bounded and uniformly Lipschitz.

By definition a solution to FBSDE (2.15), (2.16) is a triple of $\mathcal{F}_t$-measurable stochastic processes $(\gamma(t), Y(t), Z(t))$ possessing the following properties

$$\sup_{t\in[0,T]} E\|\gamma(t)\|^2 < \infty, \quad \sup_{t\in[0,T]} E\|Y(t)\|_1^2 < \infty, E\int_0^T \|Z(t)\|_1^2 dt < \infty.$$

Let us state the following assertion proved in [12].

Theorem 2.1. *Assume that* **C. 2.1** *holds. Then there exists a unique solution of the FBSDE (2.15), (2.16).*

Theorem 2.2. *Assume that* **C. 2.1** *holds. Then there the solution of the FBSDE (2.15), (2.16) gives rise to a viscosity solution of the Cauchy problem (2.14) .*

Remark 2.3. We say that the system (2.1) has a viscosity solution $u(t, y)$ if $Y(0) = \Phi(t, \gamma) = h \cdot u(t, y)$ is a viscosity solution of (2.14).

Consider the system (2.4) and show that it might be considered as a scalar equation.

Remark 2.4. The system (2.4) is equivalent to a scalar equation with respect to a scalar function $v(T - t, x, m) = v_m(T - t, x) = u_m(t, x)$ and the correspondent stochastic representation of the solution to (2.4) has the form $v_m(T - t, x) = v(T - t, x, m) = E[u_0(\xi(T), \nu(T))]$, where $\xi(t)$ satisfies the SDE

$$d\xi(\theta) = a(\xi(\theta), \zeta(\theta), v(T-\theta, \xi(\theta), \nu(\theta)))d\theta + A(\xi(\theta), \zeta(\theta), v(T-\theta, \xi(\theta), \nu(\theta)))dw(\theta),$$

$$\gamma(t) = (y, h), \zeta(t) = m,$$

and $\zeta(\theta) \in V = \{1, 2, \ldots, d_1\}$ is a Markov chain with

$$P\{\zeta(\theta) = l|\zeta(t) = m\} = q_{lm}.$$

In addition one can reduce solution of this system to solution of the corresponding FBSDE [8].

3. Numerical algorithm

Let us consider the Cauchy problem for system (2.13) assuming that it has a unique continuous solution $u(t, x) \in R^d$. As it was mentioned in the previous section to construct a solution of (2.13) it is enough to construct a solution to the FBSDE (2.15), (2.16).

It was observed [9] that one can associate with an FBSDE a certain optimization problem. Following [9] we consider the FBSDE

$$d\gamma(\tau) = q(\gamma(\tau), Y(\tau), Z(\tau))d\tau + Q(\gamma(\tau), Y(\tau), Z(\tau)) \cdot dW(\tau), \quad \gamma(t) = \gamma, \quad (3.1)$$

$$dY(\tau) = -M(\gamma(\tau), Y(\tau), Z(\tau))d\tau + Q(\gamma(\tau), Y(\tau), Z(\tau))Z(\tau) \cdot dW(\tau), \quad (3.2)$$

$$Y(T) = \Phi_0(\gamma(T)).$$

and assume that

$$q : R^n \times R \times R^\theta \to R^n, \quad Q : R^n \times R \times R^\theta \to R^\theta,$$
$$M : R^n \times R \times R^\theta \to R, \quad \Phi_0 : R^n \to R$$

satisfy conditions of the existence and uniqueness theorem for solutions of FBSDEs.

Let us treat (3.1), (3.2) as a controlled SDE with $Z(t)$ being a control process and $(\gamma(t), Y(t))$ being the state process. and consider a functional

$$J(t, \gamma, Y_t, Z(\cdot)) = E[|Y(T) - \Phi_0(\gamma(T))|^2 | \gamma(t) = \gamma]. \tag{3.3}$$

as a cost functional.

Then we can state an optimal control problem $\mathcal{Q}$ as follows:

for any initial triple $(t, \gamma, Y_0) \in [0, T] \times R^n \times R$ find $Z^*(\cdot) \in H_2(R^\theta)$ such that

$$J(t, \gamma, Y_0, Z^*(\cdot)) = \inf_{\tilde{Z}(\cdot) \in H_2} J(t, \gamma, Y_t, \tilde{Z}(\cdot)) = V(t, \gamma, Y_t). \tag{3.4}$$

If FBSDE (3.1), (3.2) admits an $\mathcal{F}_t$-adapted solution $(\gamma(t), Y(t), Z(t))$, then choosing $Y_0 = h \cdot u(t, y)$ we obtain

$$J(t, \gamma, Y_0; Z(\cdot)) = E[|Y(t) - \Phi_0(\gamma(t))|^2] = V(0, \gamma, Y_0) = 0.$$

On the contrary if the problem $\mathcal{Q}$ admits an optimal triple $(\gamma^*(t), Y^*(t), Z^*(t))$ for given initial triple $(0, \gamma, Y_0)$ with

$$V(0, \gamma, Y_0) = 0, \tag{3.5}$$

then $(\gamma^*(t), Y^*(t), Z^*(t))$ is an adapted solution of FBSDE (3.1), (3.2). Hence the global solvability of the FBSDE is equivalent to to the solvability of problem $\mathcal{Q}$ at some $(0, \gamma, Y_0)$ with additional condition (3.7) to be used to determine Y_0.

We use the above observation to derive a numerical scheme which allows to obtain an approximate solution of the FBSDE (3.1), (3.2) based on the deep FBSDE theory developed recently in a number of papers [17]– [20]. Thus the deep FBSDE theory combines probabilistic representations of a solution to the Cauchy problem for a nonlinear parabolic equation with the neural network theory.

To construct numerically a viscosity solution u of the Cauchy problem (2.13) we rewrite it as the Cauchy problem $\Phi(t, y, h) = Y(t) = h \cdot u(t, y)$, where $Y(\tau) = \Phi(\tau, \gamma(\tau)) = \eta(\tau) \cdot u(\tau, \xi(\tau))$ and (γ, Y) satisfy (3.1), (3.2) . Then we reduce solution of the FBSDE (3.1), (3.2) to solution of the following control problem :

to find

$$\inf_{Y_0, Z(\cdot)} E\left[|\Phi_0(\gamma^{Y_0, Z(\cdot)}(T)) - Y^{Y_0, Z(\cdot)}(T)|^2\right] \tag{3.6}$$

such that

$$\gamma^{Y_0, Z(\cdot)}(t) = \gamma + \int_s^t q(\gamma^{Y_0, Z(\cdot)}(\tau), Y^{Y_0, Z(\cdot)}(\tau), Z(\tau)) d\tau + \tag{3.7}$$
$$+ \int_s^t Q(\gamma^{Y_0, Z(\cdot)}(\tau), Y^{Y_0, Z(\cdot)}(\tau), Z(\tau)) dw(\tau),$$

$$Y^{Y_0, Z(\cdot)}(t) = Y_0 - \int_s^t L(\gamma^{Y_0, Z(\cdot)}(\tau), Y^{Y_0, Z(\cdot)}(\tau), Z^{Y_0, Z(\cdot)}(\tau)) d\tau + \int_s^t Z(\tau) dw(\tau), \tag{3.8}$$

where $Y_0 = Y(s)$- $\mathcal{F}_0$ adapted variable valued in R and $Z(t) \in R^\theta$ is $\mathcal{F}_t$-adapted square integrable stochastic process. The couple $(Y_0, Z(\cdot))$ is a control variable of the considered control problem.

Within this framework

$$\inf_{Y_0, Z(\cdot)} E\left[|\Phi_0(\gamma^{Y_0,Z(\cdot)}(T)) - Y^{Y_0,Z(\cdot)}(T)|^2\right] = 0$$

and infimum is achieved when $\gamma^{Y_0,Z(\cdot)}(t), Y^{Y_0,Z(\cdot)}(t), Z(t))$ satisfy (3.1), (3.2).

To solve the control problem of the form (3.6)-(3.8) it comes to be very effective to apply the neural network technique. We discuss it while considering a more advanced control problem below which is similar to schemes suggested in the recent papers [19], [20].

Let us choose the whole process $Y(\cdot)$ as a control together with $Z(\cdot)$.

Then the control problem has the form :

to find

$$\inf_{\Phi(\cdot), Z(\cdot)} E\left[|\Phi_0(\gamma^{\Phi,Z}(T)) - Y^{\Phi,Z}(T)|^2 + \int_s^T |Y^{\Phi,Z}(t) - \Phi(t,\gamma(t))|^2 dt\right], \quad (3.9)$$

where

$$\gamma^{\Phi,Z}(t) = x + \int_s^t q(\gamma^{\Phi,Z}(\tau), Y^{\Phi,Z}(\tau), Z(\tau))d\tau + \quad (3.10)$$

$$+ \int_s^t Q(\gamma^{\Phi,Z}(\tau), Y^{\Phi,Z}(\tau), Z(\tau))dW(\tau),$$

$$Y^{\Phi,Z}(t) = Y_0 + \int_s^t L(\gamma^{\Phi,Z}(\tau), y^{\Phi,Z}(\tau), Z^{Y_0,Z(\cdot)}(\tau))d\tau \quad (3.11)$$

$$- \int_s^t Z(\tau) \cdot Q(\gamma^{\Phi,Z}(\tau), Y^{\Phi,Z}(\tau), Z(\tau))dW(\tau)$$

and solution

$$\inf_{\Phi(\cdot), Z(\cdot)} E\left[|\Phi_0(\gamma^{\Phi,Z}(T)) - y^{\Phi,Z}(T)|^2 + \int_s^T |Y^{\Phi,Z}(t) - \Phi(t)|^2 dt\right] = 0,$$

is achieved when $\gamma^{\Phi,Z}(t), Y^{\Phi,Z}(t), Z(t))$ satisfy (3.10), (3.11).

To solve this optimal problem effectively one can apply the neural network technique. Let us recall some notion and results from the neural network theory.

We fix input dimension $d^0 = d + d_1$ which equals the dimension of $\gamma = (y, h)$, the output dimension $d^1 = 1$ and the global number of layers $L+1$ The first layer is an input layer having d^0, the last layer is an output layer having $d^1 = 1$ and $L-1$ layers between input and output we choose for simplicity having $m_l = m, l = 1, \ldots, L-1$.

A feedforward neutral network is a function from R^{d^0} to R^{d^1} defined as a map $\mathcal{S}^\beta : R^{d^0} \to R^{d^1}$ of the form

$$\mathcal{S}^\beta(\gamma) = \Gamma_L \circ \rho \circ \Gamma_{L-1} \circ \cdots \circ \rho_l \circ \Gamma_1(\gamma) \in R^{d^1}. \quad (3.12)$$

Here $\Gamma_l, l = 1, \ldots, L$ are affine transformations of the form $\Gamma_l(z) = B_l(z) + \beta_l$ were a matrix B_l is called weight and a vector β_l is called bias. Function $\rho : R \to R$ is called activation function, it is applied component-wise to the outputs of Γ_l, that is $\rho(\gamma_1, \ldots, \gamma_m) = (\rho(\gamma_1), \ldots \rho(\gamma_m))$. All matrices Γ_l and vectors $\beta_l, l = 1, \ldots, L$ are parameters of the neutral network. We can treat them as an element β of the space R^{N_m}, where $N_m = \sum_{l=0}^{L-1} m_l(1+n_{l+1}) = d^0(1+m)+m(1+m)(L-2)+m(1+d^1)$.

For an integer $K \in \mathbf{N}$ consider a partition $t = t_0 < t_1 < \dots < t_K = T$ of the interval $[t,T]$ and define neural network $\mathcal{S}^\beta_k(\cdot)$ to approximate the function $\Phi(t_{k+1},\cdot)$.

Set the network having one input layer with dimension n, two hidden layers with dimensions $m = 1$ and one output layer with dimension 1.

For an activation function we choose

$$\rho_n(x) = (\max(x_1,0),\dots,\max(x_n,0)),\ x \in R^d,$$

and affine transformations $\Gamma^\beta_{q,l} : R^l \to R^q$ in (3.12), are chosen to have a form

$$\Gamma^{\beta,\alpha}_{q,l}(x) = \begin{pmatrix} \beta_{\alpha+1} & \beta_{\alpha+2} & \dots & \beta_{\alpha+l} \\ \beta_{\alpha+l+1} & \beta_{\alpha+l+2} & \dots & \beta_{\alpha+2l} \\ \beta_{\alpha+2l+1} & \beta_{\alpha+2l+2} & \dots & \beta_{\alpha+3l} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{\alpha+(q-1)l+1} & \beta_{\alpha+(q-1)l+2} & \dots & \beta_{\alpha+ql} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_l \end{pmatrix} + \begin{pmatrix} \beta_{\alpha+ql+1} \\ \beta_{\alpha+ql+m} \\ \beta_{\alpha+ql+3} \\ \vdots \\ \beta_{\alpha+ql+q} \end{pmatrix}, \tag{3.13}$$

where $\alpha = d^0$.

Finally, we arrive to an optimization problem which could be approximately solved by applying stochastic descent gradient (SGD) method.

To construct the required approximate solution of (3.10), (3.11) we apply the Euler-Maruyama scheme to discretize these equations setting $\bar\gamma(t_0) = \gamma, \bar Y(t_0) = Y_0$,

$$\bar\gamma(t_{k+1}) = \bar\gamma(t_k) + q(\bar\gamma(t_k), \bar\Phi(t_k), \bar Z(t_k))\Delta_k t + Q(\bar\gamma(t_k), \bar\Phi(t_k), \bar Z(t_k))\Delta_k w, \tag{3.14}$$

$$\bar Y(t_{k+1}) = \bar Y(t_k) - L(\bar\gamma(t_k), \bar Y^1(t_k), \bar Z(t_k))\Delta t + \tag{3.15}$$
$$+\bar Z^j(t_k)\Delta_k w.$$

Here we wrote a discretization of the integral form of the backward SDE rewritten in the form similar to the integral form of a forward SDE.

We set $\bar\gamma(t_0) = \xi_0, \bar Y(t_0) = (h, u_0)$,

$$\bar\Phi(t_k) = \phi_1(\bar\gamma(t_k), \bar u(t_k); \beta_k),$$
$$\bar Z(t_k) = D\phi_1(\bar\gamma(t_k), \bar\Phi(t_k); \beta_k),$$

where D denotes the automatic differentiation operator [21].

Note that solving the last equations we can obtain

$$\bar u(t_k) = \tilde\phi_1(\bar\xi(t_k; \beta_k)),$$
$$\bar Z(t_k) = D\tilde\phi_1(\bar\gamma(t_k; \beta_k)).$$

Hence, we have to construct a neural network with a multidimensional parameter $\beta = (y_0, z(\cdot))$. The loss function has the form

$$\mathcal{L}(\beta) = \inf_\beta E\left[|\Phi_0(\bar\gamma^\beta(T)) - \bar Y^\beta(T)|^2 + \sum_{k=0}^{N-1} |\bar Y^\beta(t_k) - \bar\Phi(t_k)|^2\right].$$

Similar to [19] we may write the correspondent algorithm as follows.

Choosing as an input the Wiener process increments $\Delta_k w$, initial parameters β^0 and learning rate ϱ we have to obtain as an output the couple $(\bar{\xi}^q(T), \bar{y}^q(t_k)), k = 1\dots, N$. On each interval $[t_k, t_{k+1}]$ we solve the optimization problem applying SGD method (stochastic decent gradient) with q iterations, $q = 1, 2, \dots$.

1. For $q = 1, \dots,$ set $L = 0, \bar{\gamma}^q(t_0) = x$, $\bar{Y}^q(t_0) = \Phi(\gamma; \beta_0^{q-1})$;
2. For $k = 0, \dots, K-1$ set

$$\bar{\Phi}^q(t_k) = \phi^1(\bar{\gamma}^q(t_k), \beta_k^{q-1}),$$

$$\bar{z}^q(t_k) = D\phi(\bar{\gamma}^q(t_k), \beta_k^{q-1}).$$

3. By Euler -Maruyama schemes (3.14), (3.15) we calculate $\xi^q(t_{k+1}), y^q(t_{k+1})$ and $z^q(t_{k+1})$ on each time interval $[t_k, t_{k+1}]$

$$\bar{\gamma}^q(t_{k+1}) = \bar{\gamma}^q(t_k) + q(\bar{\gamma}^q(t_k), \bar{\Phi}^q(t_k), \bar{D}\Phi^q(t_k), \bar{z}^q(t_k))\Delta_k t +$$
$$+ Q(\bar{\gamma}^q(t_k), \bar{\Phi}^q(t_k), \bar{D}\Phi^q(t_k), \bar{z}^q(t_k))\Delta_k w,$$
$$Y^q(t_{k+1}) = Y^q(t_k) + L(\bar{\gamma}^q(t_k), \bar{Y}^q(t_k), Z^q(t_k))\Delta t -$$
$$- Z^q(t_k)\Delta_k w.$$

4.

$$(\beta^{q+1}, \bar{y}_0^{q+1}) = (\beta^q, \bar{y}_0^q) - \varrho\nabla\frac{1}{M}\sum_{m=1}^{M}\|\bar{Y}^q(T) - \Phi_0(\bar{\gamma}^q(T))\|^2.$$

4. Stochastic models of the forward Cauchy problem for type II systems

In this section we consider the forward Cauchy problem for systems which do not admit the above mentioned reduction to the backward Cauchy problem. The solutions of thses problems should be either measures or measure densities with respect to the Lebesgue measure.

Let $\mathcal{P}_2(R^d)$ is the space of probability measures with finite moments of order 2 with the 2-Wasserstein metric

$$W_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{R^d \times R^d} \|x - y\|^2 \pi(dx, dy) \right)^{\frac{1}{2}}.$$

Here $\Pi(\mu, \nu)$ is the set of measures on $R^d \times R^d$ with marginals μ and ν.

Stochastic approach to systems of the form

$$\frac{\partial u_m}{\partial t} = \frac{1}{2} Tr \nabla^2 [B_m(y, u)u_m] - div[a_m(y, u)u_m] + c_m(y, u)u_m, \quad u_m(0, y) = u_{0m}(y), \tag{4.1}$$

where $B_{ij}^m(y, u) = \sum_{k=1}^d A_{ik}^m(y, u) A_{kj}^m(y, u)$ was developed in papers [23], [24].

Here we apply a similar approach to systems of the form

$$\frac{\partial u_m}{\partial t} = \frac{1}{2} Tr \sum_{q=1}^{d_1} \nabla[B_{mq}(y, u)\nabla u_q] + c_m(y, u)u_m, \tag{4.2}$$

$$u_m(0, y) = u_{0m}(y).$$

This type of systems arises in description of various phenomena in chemistry, biology and other fields. Among systems of this type there are systems that often used to model chemotaxis processes [25] such as the well known Keller-Segel system

$$\frac{\partial n}{\partial t} = \frac{1}{2}\nabla \cdot [k_1(n,S)\nabla n - k_2(n,S)]\nabla S + H(n,S), \quad n(0,y) = n_0(y). \tag{4.3}$$

$$\frac{\partial S}{\partial t} = \frac{1}{2}k_3 \Delta S + K(n,S) \quad S(0,y) = S_0(y) \tag{4.4}$$

Here $n = n(t,y)$ is a density of cells (or organisms) interacting with a chemoattractant whose density is $S = S(t,y)$. In addition, H and K model the source terms related to interactions. A probabilistic approach to this particular case of (4.3) was developed in [22], [26].

To construct Markov processes $\xi_m(t)$ associated with the Cauchy problem (4.3) we look for their generators $\mathcal{A}_m$ defined on twice differentiable functions by a relation

$$\mathcal{A}_m\phi(x) = \lim_{s\to 0}\frac{E\phi(\xi_m(s)) - \phi(x)}{s}.$$

To obtain an explicit expression for $\mathcal{A}$ we multiply (4.3) by a test function $\phi \in C_0^\infty(R^d)$, integrate the product over R^d and apply the integration by part formula. As a result we get

$$-\frac{\partial}{\partial t}\int_{R^d}\phi(y)u_m(t,y)dy = \int_{R^d}\frac{1}{2}TrB_m(y,u(t,y))\nabla^2\phi(y)u_m(t,y)dy+ \tag{4.5}$$

$$+\frac{1}{2}\int_{R^d}\sum_{i,j=1}^{d}\nabla_i B_m^{ij}(y,u(t,y))\nabla_{y_j}\phi(y)u_m(t,y)dy+\int_{R^d}c_m(y,u(t,y))\phi(y)u_m(t,y)dy,$$

where $m \in \{1,\dots,d_1\}$.

Assume that $B_m^{ij}(y,u)$ is a positive matrix, $B_m^{ij}(y,u) = \sum_{k=1}^d A_m^{ik}(y,u)A_m^{kj}(y,u)$ for each $m = 1,\dots,d_1$ and $a_m^j(y,u(t,y)) = \frac{1}{2}\sum_{i=1}^d u_q(t,y)\nabla_{y_i}B_{mq}^{ij}(y)$.

From (4.5) we deduce that the required Markov processes $\xi_m(t)$ have generators of the form

$$\mathcal{A}_m\phi(y) = \frac{1}{2}TrB_m(y,u)\nabla^2\phi(y) + a_m(y,u)\cdot\nabla\phi(y) \tag{4.6}$$

defined on twice differential functions.

To construct the Markov processes $\xi_m(t)$ associated with (4.3) we consider an SDE

$$d\xi_m(\tau) = a_m(\xi_m(\tau),u(\tau,\xi_m(\tau)))d\tau + A_m(\xi_m(\tau),u(\tau,\xi_m(\tau)))dw(\tau),\ \xi_m(t) = \xi_{m0}, \tag{4.7}$$

where ξ_{m0} is a random variable with a distribution $P\{\xi_0 \in dx\} = \mu_0(dx)$.

Keeping in mind (4.7) we obtain

$$-\frac{\partial}{\partial t}\int_{R^d}\phi_m(y)u_m(t,y)dy = \int_{R^d}\frac{1}{2}B_m^{ij}(y,u(t,y))\nabla^2_{y_iy_j}\phi(y)u_m(t,y)dy+ \tag{4.8}$$

$$+\int_{R^d}\left[a_m^i(y,u(t,y))\nabla_{y_i}\phi_m + c_m(y,u(t,y))\phi(y)\right]u_m(t,y)dy$$

that coincides with (4.5).

Hence Markov processes $\xi_m(\tau)$ associated with (4.3) should satisfy (4.7).

Remark 4.1. In a linear case when coefficients $A(y,\mu) = A(y)$ the SDE (4.7) is called the Stratonovich form of the Ito equation. and could be rewritten in the form

$$d\xi_m(\tau) = A_m(\xi_m(\tau)) \circ dw(\tau), \quad \xi_m(s) = x, \tag{4.9}$$

where

$$[A_m(\xi_m(\tau)) \circ dw(\tau)]^i =$$

$$\sum_{k=1}^{d} A_m^{ik}(\xi_m(\tau))dw_k(\tau) + \frac{1}{2}\sum_{k,l=1}^{d} A_m^{lk}(\xi_m(\tau))]\nabla_{y_l} A_m^{ki}(\xi_m(\tau))d\tau.$$

In nonlinear case (4.9) includes unknown variables $\xi_m(\tau)$ and $u_m(t,y)$ and we need more relations to make the system closed. To derive a closing relation we assume that $u_m(t,y)$ is a density of a distribution $\mu_m(t,dy) = P\{\xi(t) \in dy\}$ that is $u_m(t,y) = \frac{d\mu_m(t,dy)}{dy}$ and consider a relation

$$\int_{R^d} \phi(y)\mu_m(t,dy) = E\left[\phi(\xi_m(t))e^{\int_0^t c_m(\xi_m(s),u_q(s,\xi_m(s)))}\right] \tag{4.10}$$

where ϕ is a function from $C_b(R^d)\hat{C}_0^\infty(R^d)$. The right hand side of (4.10) is a linear functional for any $t \in [0,T]$ for some T. If it is bounded then by the Riesz theorem it defines a measure μ.

In what follows we consider another version of a stochastic model associated with (4.3). Namely, we consider

$$d\xi_m(\tau) = A_m(\xi_m(\tau), u(\tau,\xi_m(\tau))) \circ dw(\tau), \quad \xi_m(0) = \xi_{0m}, \tag{4.11}$$

$$u_m(t,y) = E\left[\rho(y - \xi_m(t))\exp\{\int_0^t (c_m(\xi_m(s)), u(s,\xi_m(s)))]ds\}\right] \tag{4.12}$$

where $\rho : R^d \to R$ is a mollifier in R^d.

Below we show that if there exists a solution $(\xi_m(t), u_m(t,y))$ of (4.11), (4.12) and $u_m(t,y) = \int_{R^d} \rho(y-x)v_m(t,x)dx$ then $v_m(t,x)$ satisfies the Cauchy problem.

$$\frac{\partial v_m}{\partial t} = \frac{1}{2}T\nabla(B_m(y,\rho * v_m)\nabla v_m) + c_m(y,v)v_m, \tag{4.13}$$

$$v(0,y) = v_0.$$

If we set $\rho^\epsilon(y) = \epsilon^{-d}\rho(\frac{y}{\epsilon})$ where the sequence ρ^ϵ weakly converges to the Dirac measure at zero we may consider a couple $(\xi^\epsilon(t), v^\epsilon(t,y))$ solving (2.19) with $\rho(\cdot)$ replaced by $\rho^\epsilon(\cdot)$. Formally if ρ coincides with the Dirac function we obtain that v_1, v_2 satisfy (4.3).

To study systems (4.9), (4.10) and (4.11), (4.12) we need some additional notations.

Let $C_T^d = C([0,T], R^d)$ be the space of real valued continuous functions on $[0,T]$ valued in R^d with the sup-norm $\|\cdot\|_\infty$ and ξ be the canonical process on C_T^d. Denote by $\mathcal{F}_t(C_T^d) = \sigma\{\xi(\tau), 0 \le \tau \le t\}$. Let $C_b(C_T^d)$ and $C_b(R^d)$ denote the space of bounded real valued functions on C_T^d and R^d respectively, $C_0(R^d)$ be the space of real continuous functions on R^d with compact supports. Let $\mathcal{S}(R^d) = C_0^\infty(R^d)$ be the Schwartz space and $\mathcal{S}'(R^d)$ be its dual.

Denote by $\Omega = \prod \Omega_m$, $\Omega_m = \mathcal{C}^d_T$ and $\mathcal{F}_m = \sigma\{\xi_m(s); 0 \le s \le T\}$. Let $\xi_m(t)$ be canonical processes on $(\Omega_m, \mathcal{F}_m) = C([0,T]; R^d)$ that is $\xi_m : \mathcal{C}^d_T \to \mathcal{C}^d_T$ and $\xi_m(t) = \omega_m(t), t \ge 0, \omega_m \in \mathcal{C}^d$.

For $q \ge 0$ we denote by $\mathcal{P}^q_m(\mathcal{C}^d_T)$ the set of Borel probability measures $\gamma_m, m = 1, \dots, d_1$, on $\mathcal{C}^d_T$ with finite moments of order q and set

$$\mathcal{P}_q = \mathcal{P}_q(\Omega),$$

We equip $\mathcal{P}_q(C^d_T)$ with the Wasserstein distance $d^W_t(\mu, \mu_1)$ defined by

$$d^W_T(\mu,\nu)(t) = \left[\inf_{\pi \in \Pi(\mu,\mu_1)} \left\{ \int_{C^d_T} \int_{C^d_T} \sup_{0 \le s \le T} \|\xi(s,\omega) - \xi(s,\omega_1)\|^q d\pi(\omega,\omega_1) \right\} \right]^{\frac{1}{q}}$$

for all $t \in [0,T]$ and let $\|\mu\|_W$ be the correspondent norm. Here $\Pi(\mu,\mu_1)$ denotes the set of Borel probability measures in $\mathcal{P}(C^d_T \times C^d_T)$ with marginals $\pi(d\omega, C^d_T) = \mu(d\omega)$ and $\pi(C^d_T, d\omega_1) = \mu_1(d\omega_1)$.

Given a measure γ_m on $\mathcal{C}^d_T$ consider the equations

$$u_m(t,y) = \int_{\mathcal{C}^d_T} \rho(y - \xi_m(t,\omega)) M_m(t, \xi_m(\omega), u(\xi_m(\omega)))\gamma_m(d\omega), m = 1, \dots, d_1, \tag{4.14}$$

where

$$M_m(t, \xi_m(\omega), u(\xi_m(\omega))) = \exp\{\int_0^t c_m(\xi_m(s), u_m(s, \xi_m(s)))ds\}.$$

We say that condition **C 4.1** holds if:

1. For a fixed $g \in R^{d_1}$ functions $A(y,g), \nabla_y A(y,g)$ and $\nabla_g A_m(y,g)$ are bounded on R^d in the Frobenius matrix norm;

2. $\nabla A_m(y,g)$ are Lipschitz continuous functions taking values in $R^d \otimes R^d \otimes R^{d_1}$ such that

$$\|\nabla_{y_i} A_m(y,g) - \nabla_{y_i} A_m(y_1,g_1)\|^2 \le L_A[\|y-y_1\|^2 + \|g-g_1\|^2], \quad i = 1, \dots, d$$

3. $c_m(y,g)$ are bounded and Lipschitz continuous real valued functions such that

$$|c_m(y,g) - c_m(y_1,g_1)| \le L_c[\|y-y_1\| + \|g-g_1\|] \quad m = 1, \dots, d_1.$$

$$\sup_y |c_m(y, g(y))| \le K_c.$$

We consider $\xi(t)$ as the canonical process $\xi : C^d_T \to C^d_T$ defined by $\xi(t,\omega) = \omega(t)$, $0 \le t \le T, \omega \in C^d_T$.

Let $\mathcal{C}^{d_1}$ denote a linear space of R^{d_1}- valued continuous processes $\kappa(t)$ defined on the canonical space C^d_T such that

$$\|\kappa\|_\infty = E^\gamma[\sup_t \|\kappa(t)\|] = \int_{C^d_T} \sup_t \|\kappa(t,\omega)\|\gamma(d\omega) < \infty.$$

Denote by $\|\cdot\|_{\infty,K}$ a norm equivalent to the norm $\|\cdot\|_{\infty,1}$ defined by

$$\|\kappa\|^{d_1}_{\infty,K} = E^\gamma[sup_{t \le T} e^{-Kt}\|\kappa(t)\|].$$

The spaces $(\mathcal{C}, \|\cdot\|_\infty)$ and $(\mathcal{C}, \|\cdot\|_{\infty,K})$ are Banach spaces.

Theorem 4.2. *Let condition* **C 4.1** *hold and $c_m(u)$ are bounded. Then given probability measures $\gamma_m \in \mathcal{P}(C_T^d), m = 1, \ldots, d_1$, the system (4.14) has a unique solution u_m.*

Proof. We consider an operator $T^\gamma : \mathcal{C}_T^{d_1} \to C([0,T] \times R^d; R^{d_1})$ defined by

$$T_m^\gamma(\kappa)(t,y) = \int_{\mathcal{C}_T^1} \rho(y - \xi_m(t,\omega)) M_m(t, \xi_m(\omega), \kappa_m(\omega)) \gamma_m(d\omega). \tag{4.15}$$

Besides we introduce an operator $g : u \in C([0,T] \times R^d; R^{d_1}) \mapsto g(u) \in \mathcal{C}^{d_1}$ such that $g(u)(t,\omega) = u(t, \omega(t))$. As a result we have $g \circ T^\gamma$ maps $\mathcal{C}^{d_1} \to \mathcal{C}^{d_1}$. Hence (4.15) is equivalent to

$$u = (T^\gamma \circ g)(u). \tag{4.16}$$

We start with an assumption that the map $g \circ T^\gamma$ has a fixed point $\lambda \in \mathcal{C}^{d_1}$ and choose $q^\gamma = T^\gamma(\lambda)$. Since λ is the fixed point of $g \circ T^\gamma$ we get

$$\lambda = g(T^\gamma(\lambda)). \tag{4.17}$$

Thus q^γ satisfies (4.16). To prove uniqueness of the solution to (4.16) we assume the contrary. Let there exists two functions α and β satisfying (4.16), that is $\alpha = (T^\gamma \circ g)(\alpha)$ and $\beta = (T^\gamma \circ g)(\beta)$. Let $L = g(T^\gamma(X)) = g(v)$, $N = g(T^\gamma(Y))$. Since L and N are fixed points of $g \circ T^\gamma$ then the equality $L = N$ holds γ- almost everywhere. It remains to prove the map $g \circ T^\gamma$ has a unique fixed point $M \in \mathcal{C}^{d_1}$.

Given a pair $(M, N) \in \mathcal{C}^{d_1} \times \mathcal{C}^{d_1}$ for any pair $(t, y) \in [0,T] \times R^d$ we obtain an estimate

$$|T_m^\gamma(M) - T_m^\gamma(N)|(t,y) =$$

$$|\int_{\mathcal{C}^1} \rho(y - \xi_m(t,\omega))[M_m(\xi_m(\omega), L(\omega)) - M_m(\xi_m(\omega), N(\omega))]|\gamma_m(d\omega)$$

$$\leq K_0 e^{tK_c} L_c E[\int_0^t \|N(s) - L(s)|ds] \leq K_0 e^{tK_c} L_c E[\int_0^t e^{Ks} e^{-Ks} \|N(s) - L(s)|ds] \tag{4.18}$$

$$\leq K_0 e^{tK_c} L_c \int_0^t e^{Ks} E[\sup_{0 \leq \tau \leq s} e^{-\tau K} \|N(s) - L(s)\|]ds$$

$$\leq K_0 e^{tK_c} L_c \frac{e^{Kt} - 1}{K} \|N - L\|_{\infty,K}^{d_1}.$$

At the next step we consider $g(T^\gamma(N))(t) = T^\gamma(N)(t, \xi_m(t))$ and $g(T^\gamma(L))(t) = T^\gamma(L)(t, \xi_m(t))$ and show that

$$E\left[\sup_{0 \leq \tau \leq T} e^{-Kt} \|g(T^\gamma(N))(t) - g(T^\gamma(L))(t)\|\right]$$

$$= E\left[\sup_{0 \leq \tau \leq T} e^{-Kt} \|T^\gamma(N)(t, \xi_m(t)) - T^\gamma(L)(t, \xi_m(t))\|\right]$$

$$\leq K_0 e^{tK_c} L_c \frac{1}{K} \|N - L\|_{\infty,K}^{d_1}.$$

Evaluating sup over time interval $[0,T]$ and summing over m we obtain the estimate

$$\|T^\gamma(N)-T^\gamma(L)\|\le\sum_{m=1}^{d_1}\|T^\gamma_m(N)-T^\gamma_m(L)\|\le d_1K_0e^{tK_c}L_c\frac{1}{K}\|N-L\|_\infty.$$

As a result we may choose K large enough namely $K>d_1K_0e^{tK_c}L_c$ to ensure that $g\circ T^\gamma$ is a contraction map in $\mathcal{C}^{d_1},\|\cdot\|^K_\infty$ Hence by the fixed point theorem we obtain that there exists a unique solution of the equation (5.4) . □

We have proved that for a fixed measure $\gamma=(\gamma_1,\dots,\gamma_{d_1})$ there exists a unique solution u^γ of (4.16). It remains to study dependence of this solution on γ.

Lemma 4.3. *Let* **C 5.1** *holds. Then for the solution of (4.14) an estimate*

$$\|u^{\gamma_m}_m(t,y)-u^{\tilde\gamma_m}_m(t,\tilde y)\|^2\le L\|y-\tilde y\|^2+W^2_t(\gamma_m,\tilde\gamma_m),\quad m=1,\dots,d_1 \tag{4.19}$$

for any couple $(\gamma_m,\tilde\gamma_m)\in\mathcal{P}_2(C^d_T)\times\mathcal{P}_2(C^d_T)$ for all $(t,y,\tilde y)\in[0,T]\times T^r\times R^d$.

Proof. Given $(\gamma_m,\tilde\gamma_m)\in\mathcal{P}_2(C^d_T)\times\mathcal{P}_2(C^d_T)$ we derive an estimate

$$\|u^{\gamma_m}_m(t,y)-u^{\tilde\gamma_m}_m(t,\tilde y)\|^2\le 2[\|u^{\gamma_m}_m(t,y)-u^{\gamma_m}_m(t,\tilde y)\|^2+\|u^{\gamma_m}_m(t,\tilde y)-u^{\tilde\gamma_m}_m(t,\tilde y)\|^2]. \tag{4.20}$$

To verify that $u^{\gamma_m}_m(t,y)$ is Lipschitz continuous in y we note that the mollifier ρ in (4.14) is a smooth function and stochastic process $M_m(t,\xi_m(\omega),u^\gamma_m(\omega))$ is bounded. Then we have

$$\|u^{\gamma_m}_m(t,\tilde y)-u^{\tilde\gamma_m}_m\gamma_m(t,y)\|^2= \tag{4.21}$$

$$=\int_{\mathcal{C}^d_T}|\rho(y-\xi_m(t,\omega))-\rho(\tilde y-\xi_m(t,\omega)|M_m(t,\xi_m(\omega),u^{\gamma_m}(\xi_m(\omega)))\gamma_m(d\omega)$$

$$\le L_\rho e^{tK_\rho}\|y-\tilde y\|.$$

To estimate the second term in the right hand side of (4.20) we evaluate a difference

$$\|u^{\gamma_m}_m(t,\tilde y)-u^{\tilde\gamma_m}_m\gamma_m(t,\tilde y)\|^2$$

$$=|\int_{\mathcal{C}^d_T}\rho(\tilde y-\xi_m(t,\omega))M_m(t,\xi_m(\omega),u^{\gamma_m}(\xi_m(\omega)))\gamma_m(d\omega)$$

$$-\int_{\mathcal{C}^d_T}\rho(\tilde y-\xi_m(t,\tilde\omega))M_m(t,\xi_m(\tilde\omega),u^{\tilde\gamma_m}(\xi_m(\tilde\omega)))\tilde\gamma_m(d\omega)|^2$$

$$\le\int_{\mathcal{C}^d_T\times\mathcal{C}^d_T}|\rho(\tilde y-\xi_m(t,\omega))M_m(t,\xi_m(\omega),u^{\gamma_m}(\xi_m(\omega)))$$

$$-\rho(\tilde y-\xi_m(t,\tilde\omega))M_m(t,\xi_m(\tilde\omega),u^{\tilde\gamma_m}(\xi_m(\tilde\omega)))|^2\pi(d\omega,d\tilde\omega)$$

for any $\pi\in\Pi(\gamma,\tilde\gamma)$. Using the Lipschitz property of ρ, properties of the exponential and the Gronwall lemma we deduce the estimate (4.19). □

Consider a system of SDEs

$$\xi_m(t)=\xi_{0m}+\int_0^t a_m(\xi_m(\tau),u^\gamma(\tau,\xi_m(\tau)))d\tau+\int_0^t A_m(\xi_m(\tau),u^\gamma(\tau,\xi_m(\tau)))dw(\tau), \tag{4.22}$$

where $\gamma=\prod_{m=1}^{d_1}\gamma_m$ and $\gamma_m=\mathcal{L}(\xi_m)$ is the law of the process $\xi_m(t)$.

Theorem 4.4. *Assume that* **4.1** *holds. Then there exists a unique solution of the system (4.22), (4.14).*

Proof. Let us fix γ which is a product of $\gamma_m \in \mathcal{P}_2(\mathcal{C}_T^d)$. We deduce from lemma 4.3 and **C 4.1** that there exists a unique strong solution $\xi_m(t)$, $m = 1, \dots, d_1$ to (4.22). Based on the Jensen and Burkholder-Davies -Gundy inequalities we can deduce that there exists $K_0 > 0$ such that $E\sup_{0\le t\le T}\|\xi_m(t)\|^2 \le K_0[1+E\|\xi_{0m}\|^2]$. Hence the law $Q_m(\gamma_m) = \mathcal{L}(\xi_m)$ of the process $\xi_m(t)$ belongs to $\mathcal{P}_2(\mathcal{C}_T^d)$.

Consider the map $Q : \mathcal{P}_2(\mathcal{C}_T^d) \to \mathcal{P}_2(\mathcal{C}_T^d)$ and prove that it is a contractive mapping in Wasserstein metric.

Let γ_m and $\tilde{\gamma}_m$ belong to $\mathcal{P}_2(\mathcal{C}_T^d)$ while u and $\tilde{u}$ be solutions of (4.14) corresponding to γ_m and $\tilde{\gamma}_m$ respectively. Let ξ_m and $\tilde{\xi}_m$ be solutions of (4.22) corresponding to γ_m and $\tilde{\gamma}_m$ as well.

By definition of the Wasserstein metric we have

$$[d_T^W(Q(\gamma), Q(\tilde{\gamma}))]^2 \le E\sup_{0\le t\le T}\|\xi(t) - \tilde{\xi}(t)\|^2. \tag{4.23}$$

and from (4.11) and (4.19) we deduce

$$E\sup_{0\le t\le T}\|\xi(t)-\tilde{\xi}(t)\|^2 \le C[\int_0^T E\sup_{0\le s\le t}\|\xi(s)-\tilde{\xi}(s)\|^2 dt + \int_0^T [d_t^W(Q(\gamma), Q(\tilde{\gamma}))]^2 dt.$$

Finally, by the Gronwall lemma we obtain

$$E\sup_{0\le t\le T}\|\xi(t) - \tilde{\xi}(t)\|^2 \le Ce^{CT}\int_0^T [d_t^W(Q(\gamma), Q(\tilde{\gamma}))]^2 dt$$

and keeping in mind (4.23) we get

$$[d_T^W(Q(\gamma), Q(\tilde{\gamma}))]^2 \le Ce^{CT}\int_0^T [d_t^W(Q(\gamma), Q(\tilde{\gamma}))]^2 dt.$$

This estimate allows to apply the arguments of the fixed point theorem to end the proof. □

At the end we discuss connections between solutions of systems (4.11), (4.12) and (4.7), (4.10) and their conections with Cauchy problem (4.2).

5. SDEs and the Cauchy problem for PDEs

Now we discuss connections between systems (4.11), (4.12) and (4.7), (4.10). To this end we denote by $[\rho * \mu_m](t,y) = \int_{R^d}\rho(y-x)\mu_m(t,dx)$ and consider systems

$$\xi_m(t) = \xi_{0m} + \int_0^t a_m(\xi_m(\tau), u^\gamma(\tau, \xi_m(\tau)))d\tau + \tag{5.1}$$

$$\int_0^t A_m(\xi_m(\tau), u^\gamma(\tau, \xi_m(\tau)))dw(\tau), \quad \xi_m(t) = \xi_{m0} \sim \mu_0,$$

$$u^\gamma(t,y) = \int_{\mathcal{C}_T^d}\rho(y - \xi_m(t,\omega))exp\left\{\int_0^t c_m(\xi_m(s,\omega), u^\gamma(s, \xi_m(s,\omega)))\right\}\gamma_m(d\omega) \tag{5.2}$$

and

$$\xi_m(t) = \xi_{0m} + \int_0^t a_m(\xi_m(\tau), [\rho * \mu_m^\gamma](\tau, \xi_m(\tau)))d\tau + \tag{5.3}$$
$$A_m(\xi_m(\tau), [\rho * \mu_m^\gamma](\tau, \xi_m(\tau)))dw(\tau),\ \xi_m(t) = \xi_{m0} \sim \mu_0,$$

where $\mu_m(t, dy)$ is the measure defined by

$$\int_{R^d} \phi(y)\mu_m^\gamma(t, dy) = E\left[\phi(\xi_m(t))e^{\int_0^t c_m(\xi_m(s), [\rho*\mu_m^\gamma](s,\xi_m(s)))}\gamma_m(d\omega)\right] \tag{5.4}$$

for all $\phi \in \mathcal{C}_b(R^d)$ and $\mathcal{L}(\xi_m) = \gamma_m$.

Theorem 5.1. *Assume that* **C 4.1** *holds. Then, given solutions* $(\xi_m, \mu_m^\gamma), m = 1, \dots, d_1$ *of the system (5.3), (5.4), the couples* (ξ_m, u_m^γ) *satisfy (5.1), (5.2) provided* $u_m^\gamma = \rho * \mu_m^\gamma$ *and vice versa if* (ξ_m, u_m^γ) *satisfy (5.1), (5.2) then there exist measures* μ_m^γ *such that the couple* (ξ_m, μ_m^γ) *satisfy (5.3), (5.4).*

Proof.) Fix $t \in [0, T]$ and denote by $\mathcal{F}(u_m^\gamma)(t, \lambda)$ the Fourier transform of the function $u_m^\gamma(t, y)$. Let (ξ_m, u_m^γ) satisfy (5.1), (5.2) . Since $\rho \in L^1(R^d)$ we obtain from (5.2) that

$$\mathcal{F}(u_m^\gamma)(t, \lambda) = \tag{5.5}$$
$$= \mathcal{F}(\rho)(\lambda) \int_{\mathcal{C}_T^d} e^{i\lambda\cdot\xi_m(t,\omega)} exp\left\{\int_0^t c_m(\xi_m(s, \omega), u_m^\gamma(s, \xi_m(s, \omega)))\right\}\gamma_m(d\omega).$$

We deduce from the Lebesgue dominated convergence theorem that a function

$$g^\gamma(t) : \lambda \in R^d \mapsto g^\gamma(\lambda) = \int_{\mathcal{C}_T^d} e^{i\lambda\cdot\xi_m(t,\omega)} e^{\int_0^t c_m(\xi_m(s,\omega), u^\gamma(s,\xi_m(s,\omega)))}\gamma_m(d\omega) \tag{5.6}$$

is a continuous bounded function since c_m is bounded. In addition $g_m^\gamma(t)$ is non-negative definite. To verify this we consider a sequence $a_k, k = 1, \dots, d$ of complex numbers and a sequence of $y_k \in R^d, k = 1, \dots, d$. Since for all $\lambda \in R^d$ we have

$$\sum_{k=1}^d \sum_{j=1}^d a_k \bar{a}_j e^{-i\lambda\cdot(y_k - y_j)} = \left(\sum_{k=1}^d a_k e^{-i\lambda\cdot y_k}\right) \overline{\left(\sum_{j=1}^d a_j e^{-i\lambda\cdot y_j}\right)} =$$
$$|\sum_{k=1}^d \sum_{j=1}^d a_k e^{-i\lambda\cdot y_k}|^2,$$

hence $g_m^\gamma(t)$ is non-negative definite. Furthermore one can deduce from the Bochner theorem that there exists a finite nonnegative measure Borel measure $\mu_n(t)$ on R^d such that for all $\lambda \in R^d$

$$g_m^\gamma(t, \lambda) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{R^d} e^{-i\lambda\cdot y}\nu_m^\gamma(t, dy). \tag{5.7}$$

It remains to show that setting $\mu_m^\gamma(t) = \nu_m^\gamma(t)$ we obtain a measure satisfying (5.4).

Since $\nu_m^\gamma(t)$ is a finite non-negative Borel measure one can consider it as a Schwartz distribution such that $\mathcal{F}^{-1}(g_m^\gamma) = \nu_m(t)$ and for any $\phi \in C_0^\infty(R^d)$ we have $|\int_{R^d} \phi(y)\nu_m^\gamma(t, dy)| \le \|\phi\|_\infty \nu_m^\gamma(t, R^d) < \infty$.

Thus, from equalities (5.5) and (5.7) we obtain that

$$u^\gamma_m(t,y) = [\rho * \mu^\gamma_m](t,y) \tag{5.8}$$

since $\mathcal{F}(u^\gamma_m) = \mathcal{F}(\rho)\mathcal{F}(\mu^\gamma(t))$.

On the other hand setting $\langle \phi, \mu^\gamma_m(t)\rangle = \int_{R^d} \phi(y)\mu^\gamma_m(t,dy)$ we deduce applying the Fubini theorem

$$\langle \phi, \mu^\gamma_m(t)\rangle = \langle \phi, \mathcal{F}^{-1}(g_m)\rangle = \langle \mathcal{F}^{-1}(\phi), g_m\rangle = \int_{R^d} \mathcal{F}^{-1}(\phi)(\lambda)$$

$$\int_{\mathcal{C}^d_T} e^{-i\lambda\cdot\xi_m(t,\omega)} e^{\int_0^t c_m(\xi_m(s,\omega),u^\gamma(s,\xi_m(s,\omega)))}\gamma_m(d\omega)d\lambda$$

$$= \int_{\mathcal{C}^d_T}\int_{R^d} \mathcal{F}^{-1}(\phi)(\lambda)e^{-i\lambda\cdot\xi_m(t,\omega)}d\lambda e^{\int_0^t c_m(\xi_m(s,\omega),u^\gamma(s,\xi_m(s,\omega)))}\gamma_m(d\omega)$$

$$= \int_{\mathcal{C}^d_T}\left(\int_{R^d} \mathcal{F}^{-1}(\phi)(\lambda)e^{-i\lambda\cdot\xi_m(t,\omega)}d\lambda\right) e^{\int_0^t c_m(\xi_m(s,\omega),[\rho*\mu^\gamma_m](s,\xi_m(s,\omega)))}\gamma_m(d\omega)$$

$$= \int_{\mathcal{C}^d_T} \phi(\xi(t,\omega))\exp\left\{\int_0^t c_m(\xi_m(s,\omega),[\rho*\mu^\gamma_m](s,\xi_m(s,\omega)))\right\}\gamma_m(d\omega).$$

To prove the second assertion of the theorem we note that under assumption that $(\xi_m(t),\mu^\gamma_m(t))$ is a solution of (5.3), (5.4) setting $u^\gamma_m(t,y) = [\rho*\mu^\gamma_m](t,y)$ we obtain that it solves (5.1). Setting $\phi = \rho$ in (5.4) we obtain (5.2). □

Remark 5.2. The existence of a solution of the system (5.1), (5.2) is equivalent to the existence of a solution of the system (5.3), (5.4).

To prove it we observe that in the right hand side of (5.8) we get that if the Lebesgue measure of the set $\{\lambda \in R^d : \mathcal{F}(\rho)(\lambda) = 0\}$ is equal to zero, then

$$\mathcal{F}(\mu^\gamma_m(t)) = \frac{\mathcal{F}(u^\gamma_m(t,\cdot))}{\mathcal{F}(\rho)} \quad \text{a.e.} \quad t\in[0,T].$$

This shows that μ^γ_m is uniquely defined by u^γ_m and vice versa u^γ_m is uniquely defined by μ^γ_m.

Theorem 5.3. *Assume that* **C 4.1** *holds. Then the measures* μ^γ_m *defined by (5.4) satisfy the Cauchy problem*

$$\frac{\partial\mu^\gamma_m(t)}{\partial t} = \frac{1}{2}Tr\nabla[B(y,[\rho*\mu^\gamma](t))\nabla\mu^\gamma_m(t)] + c_m(y,[\rho*\mu^\gamma](t))\mu^\gamma(t), \tag{5.9}$$

$$\mu^\gamma_m(0,dy) = \mu_{0m}(dy)$$

in the weak sense, that is for every $t\in[0,T]$, $\phi\in C_0^\infty(R^d)$ *it holds*

$$\int_{R^d}\phi(y)\mu^\gamma_m(t,dy) = \int_{R^d}\phi(y)\mu_{0m}(dy)+ \tag{5.10}$$

$$+\int_0^t\int_{R^d}\phi(y)c_m(y,[\rho*\mu^\gamma_m](\tau,dy))\mu^\gamma_m(\tau,dy)d\tau$$

$$+\int_0^t\int_{R^d}\nabla\phi(y)\cdot a_m(y,[\rho*\mu^\gamma_m](\tau,dy))\mu^\gamma_m(\tau,dy)d\tau$$

$$+\frac{1}{2}\int_0^t\int_{R^d}\nabla^2\phi(y)B_m(y,[\rho*\mu^\gamma_m](\tau,dy))\mu^\gamma_m(\tau,dy)d\tau.$$

Proof. To prove (5.10) we denote by $\eta(t) = e^{\int_0^t c_m(\xi_m(s,\omega),u^\gamma(s,\xi_m(s,\omega)))ds}$ and consider the random process $\zeta(t) = \phi(\xi_m(t))\eta(t)$. Applying the Ito formula to $\zeta(t)$ and keeping in mind that $\xi_m(t)$ satisfies (5.3) we obtain

$$E\zeta(t) = E[\phi(\xi_0)] + \int_0^t E[\nabla\phi(\xi_m(\tau) \cdot a_m(\xi_m(\tau), [\rho * \mu_m^\gamma](\tau, \xi_m(\tau)))]d\tau \tag{5.11}$$

$$+\frac{1}{2}\int_0^t E[TrB_m(y, [\rho * \mu_m^\gamma](\tau, dy))\nabla^2\phi(\xi_m(\tau))]d\tau$$

$$+\int_0^t E[\phi(\xi_m(\tau)c_m(\xi_m(\tau), [\rho * \mu_m^\gamma](\tau, \xi_m(\tau)))]d\tau.$$

From the definition of $\mu_m(t, dy)$ in (5.4) and (5.11) we deduce (5.10). $\square$

References

1. Jüngel A.: *Entropy methods for diffusive partial differential equations* Springer, 2016.
2. Le D.:*Cross Diffusion Systems Dynamics, Coexistence and Persistence*, De Gruyter 2022.
3. Freidlin M.: Quasilinear parabolic equations and measures in function space, Functional Analysis and Its Applications 1 (1967) 234–240.
4. Belopolskaya Ya., Dalecky Yu. : *Investigation of the Cauchy problem with quasilinear systems with finite and infinite number of arguments by means of Markov random processes*, Izv. VUZ Mathematics, **38**, 12 (1978) 6–17.
5. Belopolskaya Ya., Dalecky Yu. : *Stochastic equations and differential geometry* Kluwer Academic Publishers, (1990).
6. Belopolskaya Ya, Woyczinski W.: *SDEs, FBSDEs and fully nonlinear parabolic systems* Rendiconti del Seminario Matematico Torino **71** 2 (2013) 209 – 219.
7. Pardoux E., Peng S.: Backward stochastic differential equations and quasilinear parabolic partial differential, *Lect. Notes in CIS* **176** (1992) 200–217.
8. E. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic pdes of second order. — Stoch. Anal. Relates Topics: The Geilo Workshop, Birkhäuser, (1996) 79–127.
9. Ma J,.Yong J.: *Forward-Backward Stochastic Differential Equations and their Applications*, Lecture Notes in Mathematics, 1702, Springer, 1999.
10. Pardoux E. Rascanu A.: *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer, 2014.
11. Zhang J.: *Backward Stochastic Differential Equations From Linear to Fully Nonlinear Theory* PTSM, **86** Springer. 2017.
12. Belopolskaya Ya.: *Probabilistic counterparts of nonlinear parabolic PDE systems Modern Stochastics and Applications* in Springer Optimization and Its Applications **90** (2014) 71–94.
13. McKean H. : *A class of Markov processes associated with nonlinear parabolic equations*, Proc. Nat. Acad. Sci. USA **59** 6, (1966) 1907–1911.
14. Carmona R, Delarue F.:*Probabilistic Theory of Mean Field Games with Applications. I* Springer 2018.
15. Belopolskaya Ya.I., Nemchenko E. I.: *Probabilistic Representations and Numerical Algorithms for Classical and Viscosity Solutions of the Cauchy Problem for Quasilinear Parabolic Systems* Journal of Mathematical Sciences **225** 5,(2017) 733–750.
16. Belopolskaya, Ya. Chubatov A. Stochastic processes associated with fully nonlinear parabolic equations arising in financial mathematics in . *ISTE Book Data Analysis and Related Applications: Recent Findings* (2023), 308-320.
17. Han J., Jentzen A., E W., *Solving High-Dimensional Partial Differential Equations Using Deep Learning*, PNAS 115 34, (2018), 8505–8510.

18. Pham H., Warin X., GermainM.: Neural networks-based backward scheme for fully nonlinear PDEs. — SN Partial Differential Equations and Applications 2 (2021), 1–21.
19. Ji S., Peng S., Peng Y., Zhang X.: *A deep learning method for solving stochastic optimal control problems driven by fully-coupled FBSDEs.* arXiv:2204.05796 (math. OC) (2022), 1–18.
20. Raissi M.: *Forward-Backward Stochastic Neural Networks:Deep Learning of High Dimensional Partial Differential Equations* arXiv preprint arXiv:1804.07010 (2018).
21. Baydin A. G., PearlmutterB. A., Radul A. A.,Siskind J. M., Automatic differentiation in machine learning: a survey, arXiv preprint arXiv:1502.05767 (2015).
22. Belopolskaya Ya, : *Stochastic models of chemotaxis processes* , Journal of Mathematical Sciences **251** 1-14.
23. Belopolskaya Ya.I. *Probabilistic interpretations of quasilinear parabolic systems.* AMS Contemporary Mathematics **734**, (2019) 39–56.
24. Carlini E., Silva F.: A Fully-Discrete Scheme for Systems of Nonlinear Fokker-Planck-Kolmogorov Equations, in: *PDE Models for Multi- Agent Phenomena* Springer. (2018), 195–218.
25. N. Bellomo, Outada N.,Soler J., Tao Y., Winkler M.: *Chemotaxis and cross-diffusion models in complex environments: Models and analytic problems toward a multiscale vision*, Mathematical Models and Methods in Applied Sciences **32,** 4, (2022) 713–792.
26. Talay D., Tomašević M.: *A new McKean–Vlasov stochastic interpretation of the parabolic–parabolic Keller–Segel model: The one-dimensional case*, Bernoulli 26(2), 2020, 1323–1353.
27. Yong J.: *Optimality Variational Principle For Controlled Forward-Backward Optimality Variational Principle For Controlled Forward-Backward Stochastic Differential Equations With Mixed Initial-Terminal Conditions*, Numerical Algebra, Control and Optimization **13**, 3 & 4 (2023), 367–391.

Yana Belopolskaya: Information technology and AI scientific centre, Sirius University of Science and Technology, Sirius, Krasnodar region, 354340, Russia.
Email address: `yana.belopolskaya@gmail.com`