On the Systems of Conservation Laws and on a New Way To Construct for them Neural Networks Algorithms

Yu. G. Rykov[*]

(Submitted by A. I. Aptekarev)

Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Moscow, 125047 Russia

Abstract—The paper is devoted to the new approach to systems of quasilinear conservation laws which leads to the alternative view of weak solution and to the possibility of developing a new type calculation algorithms for such systems on the basis of neural networks technology. The approach under consideration is the further development of variational point of view to systems of conservation laws that was earlier described by the author. In this paper the multi-dimensional setting is considered but main results are shown in one- and two-dimensional cases.

1. INTRODUCTION

The paper discusses some aspects of alternative approach to systems of conservation laws, the variational approach, that lead to the possibility of developing the appropriate theory and numerical methods in non-standard directions. Namely, the variational approach makes an accent to the notion of critical points of functionals and to the transformation of existence and uniqueness problems to the problems of functionals minimization. From the numerical point of view the minimization setting allows applying neural networks approach and constructing non-standard algorithms. These points will be discussed further in more details. The subject of hyperbolic conservation laws is a vast area where huge progress has been made in theory and numerics, and the relevant literature is also vast. Nevertheless, it occurs, first, that the general theory is still far incomplete. So, it seems desirable to widen the set of approaches that previously was used in order to build the general theory. The second point of the paper is connected with the recent explosion of artificial intelligence usage, in particular for the solution of PDEs. As it were in numerical methods, it is more efficient to use specific properties of the PDEs under consideration when constructing a numerical algorithm. So, we propose the way to develop neural networks algorithms for the systems of conservation laws that differ from usual application of this technology for PDEs. Our approach takes into account the properties of generalized solutions to quasilinear conservation laws systems and introduces non-standard form of objective function. In the present paper we only touch two points mentioned above, so we do not intend to provide any review of the literature in the area of hyperbolic conservation laws and in the theory of neural networks as well. Interested reader can have an impression on the subject by, for example, [1] for theoretical questions and [2] regarding the numerical methods; the references therein add to the picture. The modern application of neural network modeling connected with PDEs is presented, for example, in [3].

Let us take the general system of multidimensional conservation laws and consider the Cauchy problem for such system. Namely, let $(t, \mathbf{x}) \in \prod_T \equiv \{(t, \mathbf{x}) : (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m\}$, $\mathbf{U}(t, \mathbf{x}) = (u_1(t, \mathbf{x}), \ldots, u_n(t, \mathbf{x}))$, $(t, \mathbf{x}) \equiv (t, x_1, \ldots, x_m)$, and $\mathbf{F}_j = (f_{1j} \ldots, f_{nj})$, $j = 1, \ldots, m$ are sufficiently smooth (at least $\mathbf{F}_j \in C^1(\mathbb{R}^n)$) vector functions of the variables $(u_1, \ldots, u_n)$. Here and further on, vector values will be

---
[*]E-mail: yu-rykov@yandex.ru

indicated in bold in formulas. Then the Cauchy problem for the system of conservation laws reads as follows

$$D(\mathbf{U}) \equiv \frac{\partial}{\partial t}\mathbf{U}(t, \mathbf{x}) + \sum_{j=1}^{m} \frac{\partial}{\partial x_j}\mathbf{F}_j\left(\mathbf{U}(t, \mathbf{x})\right) = 0, \quad \mathbf{U}(0, \mathbf{x}) = \mathbf{U}_0(\mathbf{x}). \tag{1}$$

The solutions to system (1) that take given initial values are understood in the generalized sense with respect to the following conventional definition.

Definition 1. Let $\mathbf{U}_0(\mathbf{x}) \in \mathbb{R}^n$ be a bounded measurable function in $\mathbb{R}^m$. A bounded and measurable function $\mathbf{U}(t, \mathbf{x})$ in $\prod_T$ is called a *generalized solution* to the problem (1) if, for every test function $\varphi \in C^\infty([0, T] \times \mathbb{R}^m)$ such that $\varphi(t, \cdot) \in C_0^\infty(\mathbb{R}^m)$ for fixed $t \in [0, T]$ and $\varphi \equiv 0$ for $T_1 \leqslant t \leqslant T, T_1 < T$, the following integral identity holds

$$\iint\limits_{\prod_T} \left[\mathbf{U}\frac{\partial\varphi}{\partial t} + \sum_{j=1}^{m}\mathbf{F}_j(\mathbf{U})\frac{\partial\varphi}{\partial x_j}\right] d\mathbf{x}dt + \int\limits_{\mathbb{R}^m} \mathbf{U}_0\varphi(0, \mathbf{x})d\mathbf{x} = 0. \tag{2}$$

In case $\mathbf{U}(t, \mathbf{x})$ is continuously differentiable function the equivalence of formulations (1) and (2) is straightforward. Suppose $\mathbf{U}(t, \mathbf{x})$ is continuously differentiable function except along certain hypersurface of codimension one $\Omega \subset [0, T] \times \mathbb{R}^m$, which has continuous outward normal vector $(n_0, n_1, \ldots, n_m)$. Let $\mathbf{U}(t, \mathbf{x})$ has the discontinuity of the first kind along Ω and the values $\mathbf{U}^\pm = \mathbf{U}(t, \mathbf{x} \pm 0)$ exist. Then the relation (2) is true if the equation (1) is valid in the domain of smoothness of $\mathbf{U}(t, \mathbf{x})$ but along Ω the Rankine−Hugoniot conditions are fulfilled

$$\left(\mathbf{U}^- - \mathbf{U}^+\right)n_0 + \sum_{j=1}^{m}\left(\mathbf{F}_j(\mathbf{U}^-) - \mathbf{F}_j(\mathbf{U}^+)\right)n_j = 0. \tag{3}$$

This fact is the classical one and can be checked easily. But it is also well-known that the Definition 1 does not guarantee the uniqueness of a solution to the problem (1). Thus some additional condition is required for the function $\mathbf{U}(t, \mathbf{x})$. In the modern literature it is believed that such condition should have the form of entropy inequality.

Definition 2. Let us call convex positive function $\eta(\mathbf{U}) \in C^1(\mathbb{R}^n)$ an *entropy* for the system (1) if for the classical solutions an additional conservation law holds

$$\frac{\partial}{\partial t}\eta(\mathbf{U}(t, \mathbf{x})) + \sum_{j=1}^{m}\frac{\partial}{\partial x_j}q_j\left(\mathbf{U}(t, \mathbf{x})\right) = 0 \tag{4}$$

with some sufficiently smooth flow functions $q_j(u_1, \ldots, u_n)$.

Definition 3. The function $\mathbf{U}(t, \mathbf{x})$ which is a generalized solution to (1) in the sense of Definition 1, is called an *entropy solution* to the problem (1) if for every entropy $\eta(\mathbf{U})$ from the Definition 2 and any test function $\varphi(t, \mathbf{x}) \geq 0$ from the Definition 1 the following inequality holds

$$\iint\limits_{\prod_T} \left[\eta(\mathbf{U})\frac{\partial\varphi}{\partial t} + \sum_{j=1}^{m}q_j(\mathbf{U})\frac{\partial\varphi}{\partial x_j}\right] d\mathbf{x}dt + \int\limits_{\mathbb{R}^m} \eta(\mathbf{U}_0)\varphi(0, \mathbf{x})d\mathbf{x} \geq 0. \tag{5}$$

Again in case of piecewise continuously differentiable function $\mathbf{U}(t, \mathbf{x})$, which is an entropy solution to (1), the inequality

$$\left(\eta(\mathbf{U}^-) - \eta(\mathbf{U}^+)\right)n_0 + \sum_{j=1}^{m}\left(q_j(\mathbf{U}^-) - q_j(\mathbf{U}^+)\right)n_j \geq 0 \tag{6}$$

along a discontinuity hypersurface Ω will be true in addition to Rankine−Hugoniot relations (3).

As it is said above, the general theory in the field of quasilinear hyperbolic conservation laws systems is still far incomplete. A sufficiently complete theory has been constructed only for a single conservation

law more than fifty years ago by S.N. Kruzhkov in [4]. In the case of systems, fairly general results—A. Bressan, [5]—have been obtained only for one spatial variable and, as a rule, under the assumption that the variation range of the unknown functions is small. In multidimensional case, with some degree of conditionality, we can state that there are a plenty of partial results and no more or less general theory, consult, for example, with [6]. Thus the intention of the present paper is to widen the scope of approaches in conservation laws theory and investigate the line of thought which in a sense is alternative to the current main stream.

The paper we present is structured as follows. Section 2 highlights the main concepts of variational approach to systems of conservation laws and formulates useful properties of generalized solutions that follow from it. Section 3 describes conventional neural networks machinery and introduces new form of objective function in case of conservation laws in one-dimensional case. Finally, the possible extension to two-dimensional case is provided in Section 4.

2. VARIATIONAL POINT OF VIEW TO ONE-DIMENSIONAL SYSTEMS OF CONSERVATION LAWS

Let us first consider the one-dimensional variant of (1). Namely,

$$\frac{\partial}{\partial t}\mathbf{U} + \frac{\partial}{\partial x}\mathbf{F}(\mathbf{U}) = 0, \quad \mathbf{U}(0,x) = \mathbf{U}_0(x), \tag{7}$$

here $x \in \mathbb{R}$. The main idea of variational approach that was introduced in [7] and also in earlier publications mentioned therein, reads as follows. Consider instead of function $\mathbf{U}(t,x)$ the functional $\mathbf{J} : \chi(\tau) \in C^1\left([0,T],\mathbb{R}\right) \to \mathbb{R}^n$,

$$\mathbf{J} \equiv \int_0^T \mathbf{L}(\dot{\chi},\mathbf{U})d\tau; \quad \mathbf{L}(\dot{\chi},\mathbf{U}) \equiv \mathbf{U}\left(\tau,\chi(\tau)\right)\dot{\chi}(\tau) - \mathbf{F}\circ\mathbf{U}\left(\tau,\chi(\tau)\right). \tag{8}$$

Assume that $\mathbf{U}(t,x)$ belongs to the Oleinik's class K of piecewise continuously differentiable functions with finite number of piecewise continuously differentiable lines of discontinuity. The following theorem was proved in [7].

Theorem 1. *Let $\mathbf{U}(t,x) \in K$, and suppose that there exists a trajectory $\chi_{extr}(t) \in C^1\left([0,T],\mathbb{R}\right)$ such that $\delta\mathbf{J} = 0$ for this trajectory. Then, at the points $x = \chi_{extr}(t)$ where $\mathbf{U}(t,x) \in K$ is smooth, equations (7) hold in the classical sense, and at the points of intersection of $\chi_{extr}(t)$ with the discontinuity lines of the function $\mathbf{U}(t,x)$ the Rankine–Hugoniot relations*

$$\frac{ds}{dt}\cdot(\mathbf{U}^+ - \mathbf{U}^-) = \mathbf{F}(\mathbf{U}^+) - \mathbf{F}(\mathbf{U}^-) \tag{9}$$

are satisfied; here $x = s(t)$ is the discontinuity curve and $\mathbf{U}^{\pm} \equiv \mathbf{U}(t,s(t)\pm 0)$. Moreover, the expression for $\delta^2\mathbf{J}$ on the trajectory $x = \chi_{extr}(t)$ contains only terms depending on $(\delta\chi)^2$ (i.e. the quadratic form does not contain terms with $\delta\dot{\chi}$).

If there exist plenty enough of such extremal trajectories $x = \chi_{extr}(t)$ for given $\mathbf{U}(t,x)$ in order to cover the whole $\prod_T$ then it is easy to check that this $\mathbf{U}(t,x)$ is weak solution to (7) in the sense of Definition 1. Thus it is possible to interpret the weak solutions to (7) as such functions $\mathbf{U}(t,x)$ for which $\delta\mathbf{J} = 0$ for any trajectory $x = \chi_{extr}(t)$. This means that such $\mathbf{J}$ is in a sense "constant".

The paper [8] put this observation in more explicit form. Introduce the primitive function for $\mathbf{U}(t,x)$ with respect to variable x, i.e. $\mathbf{V}(t,x) \equiv \int^x \mathbf{U}(t,p)dp$, and consider the functional $\mathbf{I} : \chi(\tau) \in C\left([0,T],\mathbb{R}\right) \to \mathbb{R}^n$ instead of $\mathbf{J}$

$$\mathbf{I} \equiv \int_0^T \mathbf{M}\left(\mathbf{U}\right)\left(\tau,\chi(\tau)\right)d\tau \equiv \int_0^T\left[\frac{\partial\mathbf{V}}{\partial\tau}\left(\tau,\chi(\tau)\right) + \mathbf{F}\circ\mathbf{U}\left(\tau,\chi(\tau)\right)\right]d\tau. \tag{10}$$

The following theorem was stated and proved in [8].

Theorem 2. *Let $\mathbf{U}(t,x) \in K$, and suppose that there exists a trajectory $\chi_{extr}(t) \in C^1\left([0,T],\mathbb{R}\right)$ such that $\delta\mathbf{I} = 0$ for this trajectory. Then, at the points $x = \chi_{extr}(t)$ where $\mathbf{U}(t,x) \in K$ is*

smooth, equations (7) hold in the classical sense, and at the points of intersection of $\chi_{extr}(t)$ with the discontinuity lines of the function $\mathbf{U}(t,x)$ the Rankine–Hugoniot relations (9) are satisfied. Moreover, the property $\delta\mathbf{I} = 0$ means that the function $\mathbf{M}(\mathbf{U})$ is continuous across the discontinuities of $\mathbf{U}(t,x)$ and value of $\delta^2\mathbf{I}$ changes by the jump of $\partial\mathbf{M}(\mathbf{U})/\partial x$.

Let us now put Theorem 2 in the form that reflects the "constancy" of $\mathbf{J}$ mentioned above.

Theorem 3. *Suppose $\mathbf{U}(t,x) \in K$ is the generalized solution to the problem (7). Then the function $\mathbf{M}(\mathbf{U})(t,x)$, defined in (10), is continuous and does not depend on x.*

Proof. Theorem 1 demonstrates that for the functions $\mathbf{U}(t,x) \in K$ the condition $\delta\mathbf{J} = 0$ is equivalent to the fact that $\mathbf{U}(t,x)$ is the generalized solution to (7). The value of $\delta\mathbf{J}$ can be evaluated as Gateau derivative. Let fix two arbitrary trajectories $\chi_0(\tau)$ and $\chi_1(\tau)$, $\Delta\chi \equiv \chi_1 - \chi_0$, and introduce $\mathbf{J}_\alpha$ as follows

$$\mathbf{J}_\alpha \equiv \int\limits_0^T \left[\mathbf{U}(\tau, \chi_0 + \alpha\Delta\chi)(\dot\chi_0 + \alpha\Delta\dot\chi) - \mathbf{F}\circ\mathbf{U}(\tau, \chi_0 + \alpha\Delta\chi)\right]d\tau.$$

Since $\mathbf{U}(t,x) \in K$ it is enough to suppose that the function has only one continuously differentiable discontinuity line $x = s(t)$. Let $\bar\chi \equiv \chi_0(\tau) + \alpha\Delta\chi(\tau)$ and assume also that there exists one point τ_0, $\chi_0(\tau_0) = s(\tau_0)$ and one point $\tau^*(\alpha)$, $\bar\chi(\tau^*) = s(\tau^*)$. Then

$$\frac{d}{d\alpha}\mathbf{J}_\alpha = \frac{d}{d\alpha}\int\limits_0^{\tau^*(\alpha)} \mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^-\right)d\tau + \frac{d}{d\alpha}\int\limits_{\tau^*(\alpha)}^T \mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^+\right)d\tau$$

$$= \frac{d\tau^*}{d\alpha}\left[\mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^-\right) - \mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^+\right)\right] + \int\limits_0^{\tau^*(\alpha)} \frac{d}{d\alpha}\mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^-\right)d\tau + \int\limits_{\tau^*(\alpha)}^T \frac{d}{d\alpha}\mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^+\right)d\tau.$$

Further,

$$\frac{d}{d\alpha}\mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^\pm\right) = \frac{\partial\mathbf{U}}{\partial\chi}\dot{\bar\chi}\Delta\chi + \mathbf{U}\dot{\Delta\chi} - \frac{\partial\mathbf{F}\circ\mathbf{U}}{\partial\chi}\Delta\chi, \quad \frac{d\tau^*}{d\alpha} = \frac{\Delta\chi}{\dot s - \dot{\bar\chi}},$$

and, integrating by parts and taking into account that $\mathbf{U}^\pm$ satisfy (7) in classical sense, we obtain

$$\frac{d}{d\alpha}\mathbf{J}_\alpha = \frac{d\tau^*}{d\alpha}\left[\mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^-\right) - \mathbf{L}\left(\dot{\bar\chi}, \mathbf{U}^+\right)\right] + \left(\mathbf{U}^- - \mathbf{U}^+\right)\Delta\chi + \mathbf{U}\Delta\chi\big|_0^T$$

$$= \frac{\Delta\chi}{\dot s - \dot{\bar\chi}}\left[\left(\mathbf{U}^- - \mathbf{U}^+\right)\dot s - \left(\mathbf{F}\circ\mathbf{U}^- - \mathbf{F}\circ\mathbf{U}^+\right)\right] + \frac{d}{d\alpha}\int\limits_0^T \frac{d}{d\tau}\int\limits^{\bar\chi}\mathbf{U}(\tau,p)dpd\tau.$$

Thus

$$0 = \frac{d}{d\alpha}\int\limits_0^T\left[\mathbf{U}(\tau,\bar\chi)\dot{\bar\chi} - \mathbf{F}\circ\mathbf{U}(\tau,\bar\chi) - \frac{d}{d\tau}\int\limits^{\bar\chi}\mathbf{U}(\tau,p)dp\right]d\tau$$

$$= -\frac{d}{d\alpha}\int\limits_0^T\left[\mathbf{F}\circ\mathbf{U}(\tau,\bar\chi) + \frac{\partial}{\partial\tau}\int\limits^{\bar\chi}\mathbf{U}(\tau,p)dp\right]d\tau = -\frac{d}{d\alpha}\int\limits_0^T\mathbf{M}\circ\mathbf{U}(\tau,\bar\chi)d\tau$$

for any $\Delta\chi$. The continuity of $\mathbf{M}\circ\mathbf{U}$ is due to Theorem 2, hence from the last equality it follows that the function $\mathbf{M}\circ\mathbf{U}$ does not depend on $\bar\chi$. $\square$

The property of independency of $\mathbf{M}\circ\mathbf{U}(t,x)$ on x can be considered as a characterization of weak solutions. From the other side, this property could be laid down in the basis of alternative notion of generalized solution for one-dimensional systems of conservation laws. Providing this property we here only discuss the possibility to construct new types of algorithms for the generalized solutions. Namely, we need to select from some functional space/class the function $\mathbf{U}(t,x)$ with $\mathbf{U}(0,x) = \mathbf{U}_0(x)$ in any

chosen sense such that the function $\mathbf{M} \circ \mathbf{U}(t, x)$ does not depend on x for all (may be a.e.) t. To put this description in more rigorous way let us formulate important particular case for the problem (7).

Lemma 1. *Assume the values of $\mathbf{U}_0(x)$ are constants for sufficiently large $|x|$ and $\mathbf{U}(t, x) \in K$. Then the function $\mathbf{M} \circ \mathbf{U}(t, x)$ is constant for all $(t, x) \in \prod_T$.*

Proof. Because of finite speed of propagation for the generalized solutions to quasilinear conservation laws system, the function $\mathbf{U}(t, x)$ is constant in $\prod_T$ for $x < 0$, $|x|$ is sufficiently large. This means that $\mathbf{M} \circ \mathbf{U}(t, x)$ is also constant in $\prod_T$ providing the same x. For generalized solutions to (7) $\mathbf{M} \circ \mathbf{U}(t, x)$ is constant in x at least for a.e. t. But because for $x < 0$, $|x|$ is sufficiently large this constant is the same for a.e. t it will be same in the whole $\prod_T$. □

Now, take $\mathbf{U}_0(x)$ as in Lemma 1. Then in order to find the generalized solution to system (7) we need to find such $\mathbf{U}(t, x)$ that $\sup_t \left\{ V_a^b \left(\mathbf{M} \circ \mathbf{U}(t, x) \right) \right\}$ takes its minimum value. Here $V_a^b (\mathbf{W})$ denotes the variation of some function $\mathbf{W}(t, x)$ with respect to x, $x \in [a, b]$. This formulation is convenient in order to apply the calculational method with the usage of neural networks.

3. ON THE NEURAL NETWORKS ALGORITHM FOR ONE-DIMENSIONAL SYSTEMS OF CONSERVATION LAWS

As it is well known there exist plenty of numerical methods for solving the systems of quasilinear conservation laws that are based on traditional finite difference, finite volume or finite element methods, see, for example, [2]. But in the recent years a new wave of interest arises for solving such systems by artificial neural networks method (NNM), including also the calculation of irregular (for example, shocks) solutions, see [9] for further information. The main reasons for such an interest could be put, as it is mentioned in, for example, [9], as follows. First, NNM is highly compatible with the modern architecture of supercomputers and in future perspective the optimization form of the problems to systems of conservation laws is also suitable for quantum calculations. Second, NNM can naturally incorporate the big data concept which seems necessary because the amount of available experimental data and the complexity of the models increase. Third, NNM can handle the simultaneous solution over an entire parameter space; this property is very useful for calibration and validation processes.

In NNM the unknown function $\mathbf{U}(t, x)$ is represented using deep neural network. Here we only mentioned the straightforward architecture of the network for the illustration purposes. Let $\omega \equiv (t, x, \lambda_1, \ldots \lambda_p)$, where $\lambda_1, \ldots \lambda_p$ are some number of parameters that are included in the mathematical model. Then ω is the input for the feed-forward network and $\mathbf{U}$ will be the output. Suppose the network contains L hidden layers and the relations between the input and output for each component of $\mathbf{U}$ read

$$u_j^l(\omega) = b_j^l + \sum_{k=1}^{N_l} w_{jk}^l \omega_k^l(\omega), \quad \omega_k^l = \sigma \left(u_k^{l-1}(\omega) \right), \quad j = 1, \ldots, N_l, \quad l = 1, \ldots, L+1, \qquad (11)$$

where N_l is the number of neurons in the hidden layer l, $N_{L+1} = 1$, w_{jk}^l and b_j^l are the weight and bias parameters of the layer l, $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function that can be chosen according to the type of the problem under investigation. The weight and bias are the parameters that are found by the learning process, the aim of which is the minimization of objective function. The choice of objective function depends on the goal of problem solving. Usually for PDEs, in particular, the conservation laws system, the equation itself, system (7) for conservation laws, is taken as the basis expression of objective function. In contrast to this the formulation presented in Section 2 allows taking the functional

$$\sup_t \left\{ V_a^b \left(\mathbf{M} \circ \mathbf{U}(t, x) \right) \right\} \qquad (12)$$

for appropriate a, b as the objective function. The expression (12) contains one nonstandard for neural networks element—the calculation of primitive function. But this problem is already intensively addressed, see, for example, [10, 11]. The expression (12) tends to be smooth for irregular solutions as it is shown in Section 2 while the equation (7) when calculated via neural network tends to exhibit δ-shocks in the case of discontinuities. Thus objective function (12) looks preferable comparing to the usage of system (7) itself as an objective function.

One of the most popular methods of solving optimization problem in the framework of neural networks is the gradient descent often coupled with various stochastic procedures. Actually neural

network operates with the smooth functions that approximate the functions from suitable Banach space. Further in this section we consider the special class of initial data and consequently the simplification of functional (12)

$$\sup_{t,x} \mathbf{M} \circ \mathbf{U}(t,x) \tag{13}$$

in order to demonstrate other possible strategy of finding the minimum.

Theorem 4. *Assume that system (7) is strictly hyperbolic, i.e.* $\mathbf{F}'$ *has n real and distinct eigenvalues* $\varkappa_i$, $i = 1, \ldots, n$, *and full set of left eigenvectors. Let* $\boldsymbol{\Lambda}$ *is the matrix consisted of the left eigenvectors, and* $\mathbf{F} \geq 0$ *with respect to each coordinate. We also suppose that if some vector* $\mathbf{a} < 0$ *(with respect to each coordinate) then* $\boldsymbol{\Lambda}^{-1}\mathbf{a} < 0$. *Let* $\mathbf{U}_0(x) \geq 0$ *satisfies the conditions of Lemma 1 and in addition it has the finite support. Let* $\mathbf{U}(t,x) \in C^1(\prod_T)$ *and consider the set* $\mathcal{U}$ *of functions* $\mathbf{U}(t,x)$ *such that* $\mathbf{U}(0,x) = \mathbf{U}_0(x)$ *and* $\mathbf{M} \circ \mathbf{U}(t,x) \geq 0$. *Suppose that (13) attains its minimums (with respect to each coordinate) on the set* $\mathcal{U}$ *at function* $\bar{\mathbf{U}}(t,x)$. *Than these minimums equal zero and* $\bar{\mathbf{U}}(t,x)$ *is the solution to (7).*

Proof. Let first note that $\mathcal{U}$ is not empty because $\mathbf{U}_0(x)$ belongs to this set. Assume that (13) attains its minimums $\mathbf{m}$ at function $\bar{\mathbf{U}}(t,x)$. We show that if some $m_k > 0$ then there exists another function $\bar{\mathbf{U}}^*(t,x) \in \mathcal{U}$ with the property

$$\sup_{t,x} \mathbf{M} \circ \bar{\mathbf{U}}^*(t,x) < \sup_{t,x} \mathbf{M} \circ \bar{\mathbf{U}}(t,x). \tag{14}$$

Consider some increment $\delta\mathbf{U}(t,x)$, $\delta\mathbf{U}(0,x) = 0$, of function $\bar{\mathbf{U}}(t,x)$ and evaluate the difference $\Delta\mathbf{M} \equiv \mathbf{M}\left(\bar{\mathbf{U}} + \delta\mathbf{U}\right) - \mathbf{M}\left(\bar{\mathbf{U}}\right)$. Denote $\mathbf{V} \equiv \int^x \mathbf{U}(t,p)dp$ and $\Delta\mathbf{V} \equiv \int^x \delta\mathbf{U}(t,p)dp$, then we have

$$\Delta\mathbf{M} = \frac{\partial}{\partial t}\int^x \delta\mathbf{U}(t,p)dp + \mathbf{F}\left(\bar{\mathbf{U}} + \delta\mathbf{U}\right) - \mathbf{F}\left(\bar{\mathbf{U}}\right) = \frac{\partial}{\partial t}\int^x \delta\mathbf{U}(t,p)dp$$

$$+ \int_0^1 \mathbf{F}'\left(\bar{\mathbf{U}} + \lambda\delta\mathbf{U}\right)d\lambda \cdot \delta\mathbf{U} = \frac{\partial}{\partial t}\Delta\mathbf{V} + \mathbf{A}(t,x)\frac{\partial}{\partial x}\Delta\mathbf{V}, \tag{15}$$

where $\mathbf{F}'$ is the matrix of the derivative of vector function $\mathbf{F}$ and $\mathbf{A}(t,x) \equiv \int_0^1 \mathbf{F}'\left(\bar{\mathbf{U}} + \lambda\delta\mathbf{U}\right)d\lambda$.

Let us take the point $(\bar{t}, \bar{x})$ where the kth component of $\mathbf{M}\left(\bar{\mathbf{U}}\right)$ attains its supremum and consider such $\delta\mathbf{U}$ that $|\delta\mathbf{U}| \leq \varepsilon$, diam supp $\delta\mathbf{U} \leq \varepsilon$, ε is sufficiently small, and $(\bar{t}, \bar{x}) \in$ supp $\delta\mathbf{U}$. Hence from (15)

$$\Delta\mathbf{M} = \frac{\partial}{\partial t}\Delta\mathbf{V} + \mathbf{F}'\left(\bar{\mathbf{U}}(\bar{t},\bar{x})\right)\frac{\partial}{\partial x}\Delta\mathbf{V} + O(\varepsilon). \tag{16}$$

Multiplying (16) by $\boldsymbol{\Lambda}$ from the left we obtain for $i = 1, \ldots, n$

$$\mathbf{l}_i\Delta\mathbf{M} = \mathbf{l}_i\frac{d\Delta\mathbf{V}}{d\varkappa_i} + O(\varepsilon), \quad \frac{d}{d\varkappa_i} \equiv \frac{\partial}{\partial t} + \varkappa_i\frac{\partial}{\partial x}. \tag{17}$$

Now in (17) it is possible to chose the rate of change of $\Delta\mathbf{V}$ in such a way that $-\mathbf{LM}\left(\bar{\mathbf{U}}\right) < \mathbf{L}\Delta\mathbf{M} < 0$ and therefore, according to our assumptions,

$$-\mathbf{M}\left(\bar{\mathbf{U}}\right) < \Delta\mathbf{M} < 0. \tag{18}$$

First inequality in (18) tells that $\mathbf{M}\left(\bar{\mathbf{U}} + \delta\mathbf{U}\right) \in \mathcal{U}$ and the second inequality shows that $\Delta\mathbf{M}$ decreases when introducing the increment $\delta\mathbf{U}$. This contradicts the fact that the minimum with respect to the kth coordinate of (13) is attained with the function $\bar{\mathbf{U}}$. Performing the same action with respect to other coordinates of (13), if necessary, we take $\bar{\mathbf{U}}^*(t,x) = \bar{\mathbf{U}}(t,x) + \delta\mathbf{U}(t,x)$ and come to the same contradiction. Thus $\mathbf{m} = 0$. $\qquad\square$

Theorem 4 shows in simplified case that there exists direct strategy to decrease the variation of components of vector function $\mathbf{M}(\mathbf{U})$ and consequently get the generalized solution to the problem (7). This way lies in good coordination with the optimization method used by general neural networks algorithms.

4. THE APPROACH IN TWO-DIMENSIONAL CASE

A multidimensional system (1) usually is much more difficult to consider. Because of this reason, here we limit ourselves by only two-dimensional system. In order to formulate the principles of construction of neural networks algorithm similar to one presented in Section 2, we also use the variational approach as in [7]. Before the consideration of two-dimensional case in more details let us note that such variational approach can be formulated in multidimensional case as well, see [12].

Following [7] consider the functional $\mathbf{J} : S(\tau, s) \in C^1\left([0,T] \times [0,1], \mathbb{R}^2\right) \to \mathbb{R}^n$ instead of system (1) written for two-dimensional case:

$$\mathbf{J} \equiv \iint\limits_{S} \mathbf{U}dx \wedge dy + \mathbf{F}(\mathbf{U})dy \wedge dt + \mathbf{G}(\mathbf{U})dt \wedge dx. \tag{19}$$

Suppose that function $\mathbf{U}(t, x, y) \in K$ and have only one C^1 surface of discontinuity Ω. Let the surface S to be parameterized by time τ and internal parameter s, i.e. $S(\tau, s) \equiv (\chi(\tau, s), \gamma(\tau, s))$, and the surface Ω is determined by the formulas $t = \tau, x = \varphi(\tau, s), y = \psi(\tau, s)$. At this the orientation of Ω is induced by the orientation of the plane (t, x) and the positive and negative sides of Ω are determined accordingly. The corresponding values of the function $\mathbf{U}(\tau, \varphi(\tau, s), \psi(\tau, s))$ with respect to two sides of Ω will be denoted as $\mathbf{U}^{\pm}$. The next theorem was stated and proved in [7].

Theorem 5. *Let $\mathbf{U}(t, x, y) \in K$, and suppose that there exists a continuously differentiable surface $S(\tau, s)$ such that $\delta\mathbf{J} = 0$ for this surface. Then, at the points $S(\tau, s)$ where $\mathbf{U}(t, x, y)$ is smooth, equations (1) hold in the classical sense, and at the points of intersection of $S(\tau, s)$ with the discontinuity surfaces of the function $\mathbf{U}(t, x, y)$ the Rankine−Hugoniot relations (3) are satisfied.*

In order to be able discussing the algorithms of neural network type let us formulate the analog of Theorem 3 in two-dimensional case.

Theorem 6. *Suppose $\mathbf{U}(t, x, y) \in K$ is the generalized solution to the problem (1) in two-dimensional case. Let us take some C^1 functions $\varkappa(\tau, s), \nu(\tau, s)$ and denote $\mathbf{U}^1 \equiv \mathbf{U}(\tau, \omega, \nu), \mathbf{U}^2 \equiv \mathbf{U}(\tau, \varkappa, \omega)$, where ω is free variable over which integration will be performed further. Introduce two functions $\mathbf{\Phi}(\tau, s, z)$ and $\mathbf{\Psi}(\tau, s, z)$*

$$\mathbf{\Phi} \equiv \nu_s \mathbf{F}(\mathbf{U}^1) + \nu_s \frac{\partial}{\partial \tau} \int\limits^{z} \mathbf{U}^1 d\omega - \nu_\tau \frac{\partial}{\partial s} \int\limits^{z} \mathbf{U}^1 d\omega + \frac{\partial}{\partial s} \int\limits^{z} \mathbf{G}(\mathbf{U}^1) d\omega,$$

$$\mathbf{\Psi} \equiv \varkappa_s \mathbf{G}(\mathbf{U}^2) + \varkappa_s \frac{\partial}{\partial \tau} \int\limits^{z} \mathbf{U}^2 d\omega - \varkappa_\tau \frac{\partial}{\partial s} \int\limits^{z} \mathbf{U}^2 d\omega + \frac{\partial}{\partial s} \int\limits^{z} \mathbf{F}(\mathbf{U}^2) d\omega. \tag{20}$$

Then for any $\varkappa, \nu$ functions $\mathbf{\Phi}, \mathbf{\Psi}$ do not depend on z.

Proof. Taking into account the parametrization of the surface $S(\tau, s)$ put the functional (19) as follows

$$\mathbf{J} = \int\limits_{0}^{T} \int\limits_{0}^{1} \left[\mathbf{L}(\nabla\chi, \nabla\gamma, \mathbf{U})\right] ds d\tau;$$

$$\mathbf{L}(\nabla\chi, \nabla\gamma, \mathbf{U}) \equiv \mathbf{L} \equiv \mathbf{U} \cdot (\chi_s \gamma_\tau - \gamma_s \chi_\tau) + \mathbf{F}(\mathbf{U})\gamma_s - \mathbf{G}(\mathbf{U})\chi_s. \tag{21}$$

Let fix two surfaces $S_0 \equiv (\tau, \chi_0(\tau, s), \gamma_0(\tau, s))$ and $S_1 \equiv (\tau, \chi_1(\tau, s), \gamma_1(\tau, s))$, $\Delta\chi \equiv \chi_0 - \chi_1$, $\Delta\gamma \equiv \gamma_0 - \gamma_1$, and denote $\bar\chi \equiv \chi_0(\tau, s) + \alpha\Delta\chi(\tau, s), \bar\gamma \equiv \gamma_0(\tau, s) + \alpha\Delta\gamma(\tau, s)$. Also denote $\mathbf{U} \equiv \mathbf{U}(\tau, \chi(\tau, s), \gamma(\tau, s)), \overline{\mathbf{U}} \equiv \mathbf{U}(\tau, \bar\chi(\tau, s), \bar\gamma(\tau, s))$. The notaitions $\mathbf{U}^{\pm}$ or $\overline{\mathbf{U}}^{\pm}$ refer to sides of discontinuity surface and have the same meaning as $\mathbf{U}, \overline{\mathbf{U}}$. In addition we need the following notations $\mathbf{L}^{\pm} \equiv \mathbf{L}(\nabla\chi, \nabla\gamma, \mathbf{U}^{\pm})$, $\overline{\mathbf{L}}^{\pm} \equiv \mathbf{L}(\nabla\bar\chi, \nabla\bar\gamma, \overline{\mathbf{U}}^{\pm})$. Further introduce the functional $\mathbf{J}_\alpha$

$$\mathbf{J}_\alpha \equiv \int\limits_{0}^{T} \int\limits_{0}^{1} \left[\overline{\mathbf{U}} \cdot (\bar\chi_s \bar\gamma_\tau - \bar\gamma_s \bar\chi_\tau) + \mathbf{F}(\overline{\mathbf{U}})\bar\gamma_s - \mathbf{G}(\overline{\mathbf{U}})\bar\chi_s\right] ds d\tau \tag{22}$$

and by analogy to Section 3 calculate $\frac{d}{d\alpha}\mathbf{J}_\alpha$. In order to do this we need some additional geometric considerations. Let the surfaces $\overline{S} \equiv S_0 + \alpha(S_1 - S_0)$ and Ω have the single intersection line l_α, see Fig. 1. This line can be found via the relations

$$l_\alpha = (\theta, \varphi(\theta, s_2(\theta, \alpha)), \psi(\theta, s_2(\theta, \alpha))), \tag{23}$$

where the function $s_2(\theta, \alpha)$ is determined together with another function $s_1(\theta, \alpha)$ through the relations $\bar{\chi}(\theta, s_1(\theta, \alpha)) = \varphi(\theta, s_2(\theta, \alpha))$ and $\bar{\gamma}(\theta, s_1(\theta, \alpha)) = \psi(\theta, s_2(\theta, \alpha))$ with $\theta \in [\tau_1(\alpha), \tau_2(\alpha)] \subset [0, T]$. It is easy to see that the function $\tau^*(s, \alpha)$, which is inverse to the function $s = s_1(\tau^*, \alpha)$, determines the parameter τ^* of the intersection time of the line l_α and of the surface $\overline{S}$ by the plane $s = const$. Now let us calculate $\frac{d}{d\alpha}\mathbf{J}_\alpha$ denoting $I^- \equiv [0, \tau^*(s, \alpha)]$, $I^+ \equiv [\tau^*(s, \alpha), T]$,

$$\frac{d}{d\alpha}\mathbf{J}_\alpha = \frac{d}{d\alpha}\left[\int_0^1 ds \int_0^{\tau^*(s,\alpha)} \overline{\mathbf{L}}^- d\tau + \int_0^1 ds \int_{\tau^*(s,\alpha)}^T \overline{\mathbf{L}}^+ d\tau\right]$$

$$= \int_0^1 (\tau^*)_\alpha(\overline{\mathbf{L}}^- - \overline{\mathbf{L}}^+)\big|_{\tau=\tau^*(s,\alpha)} ds + \int_0^1 ds \int_{I^-} \overline{\mathbf{L}}_\alpha^- d\tau + \int_0^1 ds \int_{I^+} \overline{\mathbf{L}}_\alpha^+ d\tau$$

$$= \sum_{i=\pm}(-i)\int_0^1 (\tau^*)_\alpha\left[\overline{\mathbf{U}}^i(\bar{\chi}_s\bar{\gamma}_\tau - \bar{\gamma}_s\bar{\chi}_\tau) + \mathbf{F}(\overline{\mathbf{U}}^i)\bar{\gamma}_s - \mathbf{G}(\overline{\mathbf{U}}^i)\bar{\chi}_s\right]\big|_{\tau=\tau^*} ds$$

$$+ \sum_{i=\pm}(-i)\int_0^1 \left[\overline{\mathbf{U}}^i(\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi) - (\tau^*)_s(\overline{\mathbf{U}}^i(\bar{\gamma}_\tau\Delta\chi - \bar{\chi}_\tau\Delta\gamma) + \mathbf{F}(\overline{\mathbf{U}}^i)\Delta\gamma - \mathbf{G}(\overline{\mathbf{U}}^i)\Delta\chi)\right]\big|_{\tau=\tau^*} ds$$

$$- \sum_{i=\pm}\int_0^1 ds \int_{I^i} D(\overline{\mathbf{U}}^i)(\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi)d\tau ds$$

$$+ \sum_{i=\pm}(-i)\int_{I^i} \left[\overline{\mathbf{U}}^i(\bar{\gamma}_\tau\Delta\chi - \bar{\chi}_\tau\Delta\gamma) + \mathbf{F}(\overline{\mathbf{U}}^i)\Delta\gamma - \mathbf{G}(\overline{\mathbf{U}}^i)\Delta\chi\right]\big|_{s=0}^{s=1} d\tau + \int_0^1 \overline{\mathbf{U}}^\pm(\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi)\big|_{\tau=0}^{\tau=T} ds$$

$$= \sum_{i=\pm}(-i)\int_0^1 \left[\overline{\mathbf{U}}^i(\psi_s\varphi_\tau - \varphi_s\psi_\tau) - \mathbf{F}(\overline{\mathbf{U}}^i)\psi_s + \mathbf{G}(\overline{\mathbf{U}}^i)\varphi_s\right](\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi)\big|_{\tau=\tau^*} ds$$

$$- \sum_{i=\pm}\int_0^1 ds \int_{I^i} D(\overline{\mathbf{U}}^i)(\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi)d\tau ds$$

$$+ \sum_{i=\pm}(-i)\int_{I^i} \left[\overline{\mathbf{U}}^i(\bar{\gamma}_\tau\Delta\chi - \bar{\chi}_\tau\Delta\gamma) + \mathbf{F}(\overline{\mathbf{U}}^i)\Delta\gamma - \mathbf{G}(\overline{\mathbf{U}}^i)\Delta\chi\right]\big|_{s=0}^{s=1} d\tau + \int_0^1 \overline{\mathbf{U}}^\pm(\bar{\chi}_s\Delta\gamma - \bar{\gamma}_s\Delta\chi)\big|_{\tau=0}^{\tau=T} ds$$

$$= \frac{d}{d\alpha}\int_0^T \int_0^1 \left\{\frac{d}{d\tau}\int \left[\overline{\mathbf{U}}(\bar{\chi}_s\bar{\gamma}_\alpha - \bar{\gamma}_s\bar{\chi}_\alpha)\right] d\alpha\right.$$

$$\left. + \frac{d}{ds}\int \left[\overline{\mathbf{U}}(\bar{\gamma}_\tau\bar{\chi}_\alpha - \bar{\chi}_\tau\bar{\gamma}_\alpha) + \mathbf{F}(\overline{\mathbf{U}})\bar{\gamma}_\alpha - \mathbf{G}(\overline{\mathbf{U}})\bar{\chi}_\alpha\right] d\alpha\right\} ds d\tau$$

because $\mathbf{U}$ is the generalized solution to (1) and the Rankine–Hugoniot conditions (3) are fulfilled.

Let first $\Delta\gamma = 0$, then from the last obtained equality it follows that $\frac{d}{d\alpha}\mathbf{\Phi} = 0$, i.e. function $\mathbf{\Phi}$ does
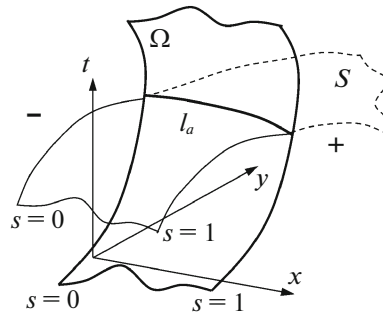
Fig. 1. Mutual disposition of the surfaces S and Ω.

not depend on α for any $\Delta\chi$ and hence on $\bar{\chi}$. The analogous statement is true concerning the function $\mathbf{\Psi}$ providing $\Delta\chi = 0$. Thus the assertion of the Theorem 6 is proved taking $\varkappa = \chi_0, \nu = \gamma_0$. $\qquad\square$

The expressions (20) are the analogy of the function $\mathbf{M}(\mathbf{U})$ in one-dimensional case. The functional (21) is the analogy of (8), expression (22) is the form of one-dimensional $\mathbf{J}_\alpha$ in two-dimensional case and the appearance of variety (line in considered case) (23) is the feature that distinguishes the multidimensional case from one-dimensional setting. Now we can introduce the equivalent of norm (12) for both functions Φ and Ψ and apply the neural networks technique described in Section 3. The new element here is the necessity to consider the representative set of functions $\varkappa(\tau, s), \nu(\tau, s)$—actually coordinate lines. Then the optimized functional should contain the norms of expressions (20) for all chosen set of coordinates. This is rather hard optimization problem, and additional research is necessary to find the way to reduce the volume of calculations.

FUNDING

REFERENCES

1. C. M. Dafermos, *Conservation Laws in Continuum Physics,* Vol. 325 of *Grundlehren der Mathematischen Wissenschaften* (Springer, Berlin, 2016).
2. J. S. Hesthaven, *Numerical Methods for Conservation Laws: From Analysis to Algorithms* (SIAM, Philadelphia, 2018).
3. D. Tarkhov and A. Vasilyev, *Semi-Empirical Neural Network Modeling and Digital Twins Development* (Academic, Elsevier, 2020).
4. S. N. Kruzhkov, "First order quasilinear equations in several independent variables," Mat. USSR Sb. **10**, 217−243 (1970).
5. A. Bressan, *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem* (Oxford Univ. Press, New York, 2000).
6. P. D. Lax, *Hyperbolic Partial Differential Equations,* Vol. 14 of *Courant Lecture Notes in Math.* (AMS, Providence, RI, 2006).
7. Yu. G. Rykov, "On the variational approach to system of quasilinear conservation laws," Proc. Steklov Inst. Math. **301**, 213−227 (2018).
8. Yu. G. Rykov, "Extremal properties of the functionals connected with the systems of conservation laws," Math. Montisnigri **46**, 21−30 (2019).
9. C. Michoski, M. Milosavljevic, T. Oliver, and D. R. Hatch, "Solving differential equations using deep neural networks," Neurocomputing **399**, 193−212 (2020).
10. H. Li, Y. Li, and S. Li, "Dual neural network method for solving multiple definite integrals," Neural Comput. **31**, 208−232 (2019).
11. S. Changdar and S. Bhattacharjee, "Solution of definite integrals using functional link artificial neural network," arXiv: 1904.09656v1 (2019).
12. A. I. Aptekarev and Yu. G. Rykov, "Variational principle for multidimensional conservation laws and pressureless media," Russ. Math. Surv. **74**, 1117−1119 (2019).