Chapter 3

Generalized Solutions of Bellman's Differential Equation

3.0. Introduction

The theory of new distributions introduced in Chapter 1 can be used to define generalized solutions of the Hamilton–Jacobi–Bellman equation just as in the conventional linear theory, by using the adjoint operator. Idempotent analysis provides a physically natural interpretation of formulas thus obtained: they are given by the convolution (in the new sense) of the initial data with the Green function (the solution whose initial value is given by the idempotent δ-function), which is obtained as the solution of the related variational problem with fixed endpoints and fixed time interval.

This concept of generalized solutions originally arose from the analysis of the Burgers equation, which had earlier served as a starting point for Hopf's vanishing viscosity method. This example was considered in detail by V. P. Maslov in the book [**?**] and in the paper [**?**].

The first section of this chapter is introductory. We recall how first-order equations arise in optimal control and discuss the problem of constructing generalized solutions. Section 2 mostly deals with the general conception of weak (in the sense of idempotent analysis) solutions of evolution and stationary Hamilton–Jacobi–Bellman equations. As was shown in [**?**], this theory, with minor modifications, can be extended to more general first-order partial differential equations. Similar constructions for quasilinear equations were carried out in [**?**].

Since the differential Hamilton–Jacobi–Bellman equation is linear with respect to the operations $\oplus = \min$ and $\otimes = +$, we can look for linear finite-difference approximations to this equation. By writing out such an approximation on analogy with a general pseudodifferential operator that is a difference operator with respect to time, we obtain the general difference Bellman equation. The following question is important: in what sense do the solutions of the difference Bellman equation converge to those of the differential Bellman equation? It turns out that weak convergence (studied in detail in Chapter 1) takes place. We can use this fact in two ways: either solve the differential equation approximately with the aid of the difference equation, or obtain asymptotic formulas for solutions of discrete optimization problems with a large parameter, solutions that assume finitely many values (say, $\mathbb{0}$ and $\mathbb{1}$) and hence should seemingly have no limit at all. However, the weak limit in the cited sense (the limit in idempotent measure of open sets) may well exist. If we treat the discrete Bellman equation as an idempotent analog of Markov chains, then, on further analogy with probability theory, we can say that the theory of generalized solutions of the differential Bellman equation allows one to prove limit theorems for idempotent measures. A result of this sort can be found in §3.9.

In §3.2 we also show that to each critical point of the Hamiltonian there corresponds a generalized eigenfunction of the resolving operator of the Cauchy problem for the corresponding differential Hamilton–Jacobi–Bellman equation. Locally, this function is a generating function of the expanding Lagrangian manifold of this point. As was already noted, the eigenfunctions and eigenvalues of the Bellman operators determine the behavior of solutions to optimization problems on the infinite planning horizon.

In the case of the differential Bellman equation, these eigenfunctions have yet another application: they are limits as $t \to \infty$ of logarithmic asymptotics of solutions of tunnel-type equations. This matter is comprehensively discussed in Chapter 4. The eigenfunctions of the resolving operator of the Cauchy problem also specify generalized solutions of the stationary equation. Here we present a criterion for the semimodule of solutions of the stationary Bellman equation to be finite-dimensional and a theorem describing the limit behavior of the Cauchy problem for large time.

The remaining sections of this chapter are independent from one another and develop the ideas of §3.2 in various directions.

Section 3.3 deals with the theory of stochastic perturbations of equations linear in idempotent semimodules. These results, which were obtained in [**?**], are applied to quantum filtering methods of analysis of continuously observed quantum systems.

In §3.4 we derive and analyze a new differential Bellman type equation that describes the dynamics of Pareto sets in multicriterial optimization problems. The exposition is based on the paper [**?**]. As a result, studying the dynamics of Pareto sets in the general situation is reduced to solving a family of conventional optimization problems with a parameter. We give a simple example of how the proposed technique is applied to a variational problem with two quadratic Lagrangians.

Generalized solutions of the first boundary value problem for the stationary Hamilton–Jacobi equation $H(x, \partial S/\partial x) = 0$ are considered in §3.5. Systematic use of the idempotent superposition principle enables us to define a natural space (more precisely, an idempotent semimodule) in which the solution of the considered problem exists and is unique. Our exposition in this section follows [**?**]. Let us explain what this theory gives by considering a simple example. Consider the equation $(y'(x))^2 = 1$ on the interval $[-1, 1]$ with the boundary conditions $y(-1) = y(1) = 0$. It turns out that in an appropriately chosen semimodule of distributions corresponding to the arithmetic with the operations $\oplus = \min$ and $\otimes = +$, this problem has the unique solution $y = 1 - |x|$. This is just the solution given by the vanishing viscosity method. On the other hand, if we start from the semimodules of distributions corresponding to the arithmetic with the operations $\oplus = \max$ and $\otimes = +$, then the unique solution of the problem in question is $\tilde{y} = |x| - 1$. Note also that the natural basis in the semimodule of solutions of the equation $(y'(x))^2 = 1$ on the interval $[-1, 1]$ is formed by the functions $y_1 = x$ and $y_2 = -x$ and that the cited solution of the boundary value problem is uniquely determined as the linear combination $y = a_1 \otimes y_1 \oplus a_2 \otimes y_2$ of basis elements that satisfies the boundary conditions $y(\pm 1) = 0$.

The stochastic Hamilton–Jacobi–Bellman equation is considered in §3.6, and in §3.7 we deal with the turnpike theorem in general topological spaces and with the limit theorem for the infinite-dimensional Hamilton-Jacobi equation.

Section 3.8 outlines the idempotent counterpart of calculus of variations recently constructed by M. D. Bronshtein and some of its applications to equations of mathematical physics.

In §3.9, Bellman's differential equation is applied to studying discrete optimization problems with a large parameter.

Yet another application of idempotent analysis to partial differential equations can be found in [**?**].

3.1. First-Order Differential Equations in Calculus of Variations, Optimal Control, and Differential Games. The Problem of Constructing Generalized Solutions

In this section, we recall how first-order differential equations (such as the Hamilton–Jacobi, the Bellman, and the Isaacs equations) arise in calculus of variations, optimal control, and differential games. We also discuss the difficulties encountered in constructing generalized solutions of the Cauchy problem.

Let us start from calculus of variations. Let $S(t,x,\xi)$ be the minimum of the functional $\int_0^t L(x,\dot{x})\,dt$ over all smooth curves $x(\tau)$ joining the points $\xi = x(0)$ and $x = x(t)$; it is assumed that the function $L(x,v)$, referred to as the *Lagrangian* of the variational problem, is strictly convex with respect to v and smooth. Then for small $t>0$ the Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H\left(x, \dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.1}$$

for the Hamilton–Jacobi equation with Hamiltonian function H that is the Legendre transform of L with respect to v (that is, $H(x,p) = \max_v(pv - L(x,v))$) and with smooth initial function $S_0(x)$ has a unique solution $S(t,x)$, which is given by the formula

$$S(t,x) = (R_t S)(x) = \min_{\xi}(S(t,x,\xi) + S_0(\xi)). \tag{3.2}$$

In other words, the classical smooth solution of the Cauchy problem (3.1) is a solution of the variational minimization problem

$$\int_0^t L(x,\dot{x})\,dt + S_0(x(0)) \to \min \tag{3.3}$$

with fixed time and free left endpoint. The two-point function $S(t,x,\xi)$ itself can also be obtained from Eq. (3.2) by formally substituting the discontinuous function

$$\delta_\xi(x) = \begin{cases} 0, & x = \xi, \\ +\infty, & x \neq \xi, \end{cases} \tag{3.4}$$

(which is shown in Chapter 1 to be the idempotent analog of the usual δ-function) for $S_0(x)$.

Furthermore, it turns out that similar differential equations occur in solving optimal control problems; however, in contrast with the classical calculus of variations, the corresponding Hamiltonian is usually not smooth and the Lagrangian L is discontinuous and assumes the value $+\infty$ at some points. For example, consider the minimization problem (3.3) for a controlled object with dynamics described by the equation $\dot{x} = f(x, u)$, when u is a control parameter ranging over a compact subset U in Euclidean space. Putting it other way, the minimum in (3.3) is sought not over all curves $x(\tau)$ with endpoint $x = x(t)$, but only over the curves that satisfy the dynamic equations $\dot{x} = f(x, u(t))$ for some choice of the control $u(t) \in U$. In optimal control theory, such problems are usually represented in the equivalent form

$$\int_0^t g(x, u)\, dt + S_0(x(0)) \to \min, \tag{3.5}$$

where $g(x, u) = L(x, f(x, u))$. Obviously, problem (3.5) can also be represented in the form

$$\int_0^t \widetilde{L}(x, \dot{x})\, dt + S_0(x(0)) \to \min, \tag{3.6}$$

obtained from (3.3) by replacing L with the discontinuous Lagrangian

$$\widetilde{L}(x, \dot{x}) = \begin{cases} L(x, \dot{x}) & \text{if } \dot{x} = f(x, u) \text{ for some } u \in U, \\ +\infty & \text{otherwise.} \end{cases}$$

By formally applying the laws of the classical calculus of variations to the latter Lagrangian, we can conclude that the minimum of $S(t, x)$ in the optimal control problem (3.5) or (3.6) is a (generalized, in some sense) solution of the Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + \widetilde{H}\left(x, \dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.7}$$

for the Hamilton–Jacobi equation with nonsmooth Hamiltonian

$$\widetilde{H}(x, p) = \max_v (pv - \widetilde{L}(x, v)) = \max_{u \in U} (pf(x, u) - L(x, f(x, u))). \tag{3.8}$$

The equivalent form

$$\frac{\partial S}{\partial t} + \max_{u \in U}\left(\left\langle \frac{\partial S}{\partial t}, f(x,u)\right\rangle - g(x,u)\right) = 0 \tag{3.9}$$

of the Hamiltonian–Jacobi equation with Hamiltonian $\widetilde{H}$ is called the *nonstationary Bellman differential equation.*

Let us now recall how this equation can be obtained in optimal control theory by Bellman's dynamic programming method. Note, however, that the subsequent calculations, though intuitively very natural, are in fact heuristic. They can be given precise meaning only for smooth solutions $S(t,x)$ (or at least in a neighborhood of smoothness points), whereas S is generally nonsmooth and even discontinuous if so is the initial condition S_0.

By applying the optimality principle, according to which an optimal trajectory remains optimal after deleting an arbitrary initial segment, we can readily obtain the approximate equation

$$S(t,x) = \min_u (S(t-\Delta t, x - \Delta x(u)) + g(x,u)\Delta t). \tag{3.10}$$

Let us expand $S(t-\Delta t, x-\Delta x)$ in the Taylor series, neglect the terms of the order of $(\Delta t)^2$, and take into account the fact that $\Delta x = f(x,u)\Delta t$ in this approximation. Then we obtain

$$S(t,x) = \min_u \left\{ S(t,x) - \Delta t \left(\frac{\partial S}{\partial t}(t,x) + \left\langle f(x,u), \frac{\partial S}{\partial x}(t,x) \right\rangle - g(x,u) \right) \right\},$$

which is obviously equivalent to

$$\min_u \left(-\frac{\partial S}{\partial t}(t,x) - \left\langle f(x,u), \frac{\partial S}{\partial x}(t,x)\right\rangle + g(x,u)\right) = 0.$$

On replacing min by max and changing the sign, we obtain (3.9). Note that for the case of a single terminal payment, when $g(x,u) \equiv 0$, Eq. (3.9) is reduced to the equation

$$\frac{\partial S}{\partial t} + \max_u \left\langle \frac{\partial S}{\partial x}, f(x,u)\right\rangle = 0, \tag{3.11}$$

whose Hamiltonian is first-order homogeneous with respect to the momenta.

The above considerations suggest that the solutions of minimization problems in calculus of variations and optimal control theory are, in some sense, generalized solutions of the Cauchy problem for equations with nonsmooth Hamiltonians. The initial functions can be nonsmooth and even discontinuous. An example is given by the initial condition (3.4), satisfied by the two-point function $S(t, x, \xi)$, which is a solution of the variational problem with fixed endpoints. Discontinuous initial functions occur also in problems where the left endpoint $x(0)$ must lie on some fixed submanifold $M \subset \mathbb{R}^n$. For example, problem (3.3) with the additional condition $x(0) \in M$ can be represented in the form (3.3) with the cutoff initial function

$$\widetilde{S}_0 = \begin{cases} S_0(x), & x \in M, \\ +\infty, & x \notin M. \end{cases}$$

A distinguishing feature of the Hamiltonians (3.8) occurring in calculus of variations and optimal control theory is that they are convex (possibly, nonstrictly) in p; in particular, they are continuous in p. Hamiltonians of different form occur in differential games. Let $S(t, x_0, y_0)$ be the guaranteed approach in time t for an object P with dynamics $\dot{x} = f(x, u)$ pursuing an object E with dynamics $\dot{y} = g(y, v)$. Let $u \in U$ and $v \in V$ be the control parameters and $x_0, y_0 \in \mathbb{R}^3$ the initial states of the objects P and E, respectively. Just as above, we apply the optimality principle and find that $S(t, x, y)$ satisfies the Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + \max_u \left\langle \frac{\partial S}{\partial x}, f(x, u) \right\rangle + \min_v \left\langle \frac{\partial S}{\partial y}, g(y, v) \right\rangle = 0 \tag{3.12}$$

with the initial condition $S_0(x, y)$ equal to the distance between x and y in $\mathbb{R}^3$. Equation (3.12) with the Hamiltonian

$$H(x, y, p_x, p_y) = \max_u \langle p_x, f(x, u) \rangle + \min_v \langle p_y, g(y, v) \rangle \tag{3.13}$$

is known as the *Isaacs equation* in the theory of differential games. The Hamiltonian (3.13), in contrast with (3.9) and (3.11), is convex-concave with respect to the momenta p_x and p_y. The theory of the Cauchy problem for equations with convex-concave Hamiltonians (by the way, any Hamilton–Jacobi equation, in a sense, can be reduced to such a problem) is much more complicated than the theory of the Cauchy problem for equations with convex Hamiltonians and is not discussed in the present book. Some related results can be found in [**?**, **?**, **?**].

Stationary first-order equations of the form $H(x, \partial S/\partial x) = 0$ play an important role in optimal control theory. The simplest example is given by the classical optimal time problem [**?**]: find a control $u(t)$ such that the trajectory of the equation $\dot{x} = f(x, u)$ issuing from a given point ξ reaches a given point x in minimal time. Just as above, by applying the optimality principle, we can show that the minimal time $T(x)$ of transition from ξ to x satisfies the stationary equation

$$\max_u \langle T'(x), f(x, u) \rangle = 1 \tag{3.14}$$

at every point where $T(x)$ is smooth.

Formally, the solution $T(x)$ of the optimal time problem can be expressed via the solution of the Cauchy problem for the nonstationary Hamilton–Jacobi equation with discontinuous initial data. Namely, let us define a function $S(t, x, \xi)$ as follows: $S(t, x, \xi) = 0$ if there exists a trajectory from ξ to x in time t, and $S(t, x, \xi) = +\infty$ otherwise. Then $S(t, x, \xi)$ is a generalized solution of the Cauchy problem for Eq. (3.10) with the initial function (3.4). It readily follows from the definition of $T(x)$ that

$$T(x) = \inf_{t \geq 0} (t + S(t, x, \xi)) \tag{3.15}$$

Needless to say, to give precise meaning to the above heuristic considerations, one should correctly define the notion of a generalized solution of the Cauchy problem for the Hamilton–Jacobi equation with nonsmooth Hamiltonian convex with respect to the momenta and with discontinuous initial data; the definition should be independent of any optimization problems. Furthermore, we must require the existence and uniqueness theorem to be valid for such a definition and the solution to be in some sense continuous with respect to the initial data. It is also desirable that for Hamiltonians of the form (3.9), which occur in optimal control theory, the solution thus defined be a minimum in the corresponding optimization problem and (in some sense) the limit of solutions of the discrete Bellman equation (3.10). We now proceed to discussing the definition and construction of such generalized solutions.

So, what difficulties occur when one tries to give a reasonable definition of a solution? First, the classical (i.e., smooth) solution of the Cauchy problem (3.1) does not exist for large time even for smooth H and S_0. This phenomenon is due to the fact that the characteristics, i.e., the projections on the x-space of the solutions of Hamilton's equations

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p}(x, p), \\ \dot{p} = -\dfrac{\partial H}{\partial x}(x, p), \end{cases}$$

issuing from the points of the initial Lagrangian manifold $p = \partial S_0/\partial x$, may intersect (e.g., see [**?**]). All the more, one cannot hope to obtain smooth solutions for nonsmooth H and S_0.

The obvious idea of defining a generalized solution, at least for continuous initial data, as a continuous function that is smooth almost everywhere and satisfies the equation at the differentiability points fails immediately, since there are infinitely many such solutions even for smooth H and S_0. Indeed, consider the following simple example [**?**]. In the Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + \dfrac{1}{2}\left(\dfrac{\partial S}{\partial x}\right)^2 = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.16}$$

for the Hamilton–Jacobi equation with Hamiltonian $H(x,p) = \frac{1}{2}p^2$, $x, p \in \mathbb{R}$, let us impose the zero initial data $S_0(x) = 0$. It is easy to see that for each $\delta > 0$ the function

$$S_\delta(t,x) = \begin{cases} 0, & |x| \geq \delta t, \\ \delta|x| - \frac{1}{2}\delta^2 t, & |x| \leq \delta t, \end{cases}$$

is continuous and that outside the rays $|x| = \delta t$, $t > 0$, it is differentiable and satisfies the equation $\partial S/\partial t + \frac{1}{2}(\partial S/\partial x)^2 = 0$.

Finally, recall that, in contrast with the theory of linear equations, where generalized solutions can be defined in the standard way as functionals on the space of test functions, there is no such approach in the theory of nonlinear equations.

The most popular approach to the theory of generalized solutions of the Hamilton–Jacobi equation is the vanishing viscosity method. For the first time, this method was applied by Hopf to the Cauchy problem (3.16). Hopf suggested to define a generalized solution of problem (3.16) as the limit as $h \to 0$ of the solution to the Cauchy problem for the parabolic equation

$$\frac{\partial w}{\partial t} + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2 - \frac{h}{2}\frac{\partial^2 w}{\partial x^2} = 0 \tag{3.17}$$

with small parameter h and with the same initial condition $w_0(x,h) = S_0(x)$: $S(t,x) = \lim_{h\to 0} w(t,x,h)$. Equation (3.17) is the integrated Burgers equation. The substitution $u = \exp(-w/h)$ reduces Eq. (3.17) to the linear heat equation

$$\frac{\partial u}{\partial t} = \frac{h}{2}\frac{\partial^2 u}{\partial x^2}. \tag{3.18}$$

The solution of Eq. (3.18) with the initial condition $u_0 = \exp(-w_0/h)$ is given by the formula

$$u(t,x,h) = \frac{1}{\sqrt{2\pi h t}} \int \exp\left\{-\frac{(x-\xi)^2}{2th}\right\} u_0(\xi)\, d\xi.$$

The last expression readily implies the following formula for the solution of the Cauchy problem for Eq. (3.17):

$$\begin{aligned} w(t,x,h) &= -h \ln u(t,x,h) \\ &= -h \ln\left(\frac{1}{\sqrt{2\pi h t}} \int \exp\left\{-\frac{(x-\xi)^2}{2th} - \frac{w_0(\xi)}{h}\right\} d\xi\right). \end{aligned} \tag{3.19}$$

Thus, the definition of a generalized solution to the Cauchy problem (3.18) is reduced to calculating the logarithmic limit for the heat equation, which can readily be performed by evaluating the asymptotics of the integral in (3.19) by the Laplace method.

By extending this method to more general equations, we arrive at the definition of a generalized solution to problem (3.1) as the limit as $h \to 0$ of solutions of the Cauchy problem

$$\begin{cases} \dfrac{\partial w}{\partial t} + H\left(x, \dfrac{\partial w}{\partial x}\right) - \dfrac{h}{2}\dfrac{\partial^2 w}{\partial x^2} = 0, \\ w|_{t=0} = w_0(x) = S_0(x). \end{cases} \tag{3.20}$$

One cannot obtain a closed analytic expression for the solution $w(t, x, h)$ of this problem. However, for continuous initial data and under some reasonable restrictions on the growth of H and S_0, one can prove that there exists a unique smooth solution $w(t, x, h)$ of problem (3.20) and that the limit $S(t, x) = \lim_{h\to 0} w(t, x, h)$ exists and is continuous. Furthermore, it turns out that the solutions thus obtained are selected from the set of continuous functions satisfying the Hamilton–Jacobi equation almost everywhere by some simple conditions on the discontinuities of the derivatives. In some cases, these conditions have a natural physical interpretation. A detailed exposition of the vanishing viscosity method can be found in [**?**, **?**, **?**].

However, according to [**?**], this method cannot be used to construct a reasonable theory of generalized solutions of Eq. (3.1) for discontinuous initial functions. Furthermore, it is highly desirable to devise a theory of problems (3.1) on the basis of only intrinsic properties of Hamilton–Jacobi equations (i.e., regardless of the way in which the set of Hamilton–Jacobi equations is embedded in the set of higher-order equations). Such a theory, including solutions with discontinuous initial data, can be constructed for equations with convex Hamiltonians on the basis of idempotent analysis; in fact, the set of solutions of such equations is equipped with an idempotent semigroup superposition law, and the theory of generalized solutions can be constructed by analogy with linear equations, with the new superposition law replacing the usual one.

The form of this superposition law can be guessed easily from the simple equation (3.16) and from Hopf's method for solving this equation. Let us consider the substitution $w = -h \ln u$, which transforms the heat equation into Eq. (3.16), as a semiring isomorphism. Obviously, this transformation takes the semiring $\mathbb{R}_+$ (with standard addition and multiplication) to the semiring A_h with the "sum"

$$a \oplus b = -h \ln(\exp(-a/h) + \exp(-b/h)) \tag{3.21}$$

and the "product"

$$a \odot \lambda = a + \lambda. \tag{3.22}$$

The zero is taken to $\mathbb{0} = +\infty$ and the unity to $\mathbb{1} = 0$; these are the neutral elements with respect to addition $\oplus$ and multiplication $\odot$ in the semiring A_h.

The linear combination $u = \lambda_1 u_1 + \lambda_2 u_2$ of positive solutions to the heat equation with positive coefficients is taken to the "linear-combination"

$$w = -\ln\left(\exp\frac{\mu_1 - w_1}{h} + \exp\frac{\mu_2 - w_2}{h}\right)$$

of the solutions w_1 and w_2 to Eq. (3.17) in the semiring A_h; here $\mu_i = -\ln\lambda_i$, $i = 1, 2$. Obviously, this combination is also a solution of Eq. (3.17). Thus, we can say that Eq. (3.17) is linear with respect to the semiring operations $\oplus$ and $\odot$. Let us equip the semimodule of A_h-valued functions with the inner product

$$\langle w_1, w_2\rangle_{A_h} = -h\ln\int\exp\left(-\frac{w_1 + w_2}{h}\right)dx, \tag{3.23}$$

which is obviously bilinear:

$$\begin{aligned} \langle w_1 \oplus w_2, w_3\rangle_{A_h} &= \langle w_1, w_3\rangle_{A_h} \oplus \langle w_2, w_3\rangle_{A_h}, \\ \lambda \odot \langle w_1, w_2\rangle_{A_h} &= \langle \lambda \odot w_1, w_2\rangle_{A_h}. \end{aligned} \tag{3.24}$$

Then the resolving operator $\mathcal{L}_t w_0$ (3.19) of the Cauchy problem for Eq. (3.17) is self-adjoint with respect to (3.23) (see the book [**?**] for detailed computations).

The Burgers equation (3.17) passes as $h \to 0$ into the Hamilton-Jacobi equation (3.16); the semiring A_h passes as $h \to 0$ into the semiring $A = \mathbb{R} \cup \{+\infty\}$ with operations $a \oplus b = \min(a, b)$ and $a \odot b = a + b$. Furthermore, the inner product (3.23) passes into the inner product of A-valued functions, introduced in Chapter 1. Hence, it is natural to expect that the resolving operator of the Cauchy problem for the Hamilton-Jacobi equation is linear in the semimodule of A-valued functions, which was studied in Chapters 1 and 2. This is the case indeed. More precisely, it turns out that the classical resolving operator of problem (3.1) with convex Hamiltonian is defined and linear with respect to the operations $\oplus = \min$ and $\odot = +$ on the set of smooth convex initial functions. Since $(\oplus, \odot)$-linear combinations of such functions are dense in the semimodule $C_{\mathbf{0}}^{\infty}(\mathbb{R}^n)$, introduced in Chapter 2 (they approximate idempotent δ-functions) and since this semimodule is naturally embedded in the dual semimodule with the help of the inner product

$$\langle f, g\rangle_A = \inf_x f(x) \odot g(x), \tag{3.25}$$

it follows that the generalized solutions of (3.1) can naturally be defined as in the ususal linear theory, via the adjoint operator. This program is implemented in the next section.

Differentiation is known to reduce the Hamilton-Jacobi equation to a system of first-order quasilinear equations. The semimodule of continuous A-valued functions is then transformed into an interesting function semimodule, which is not a semimodule of semiring-valued functions and which can be used to construct generalized solutions of quasilinear equations [?]. For example, let us discuss the situation for the (nonintegrated) Burgers equation.

Consider the space of pairs $\{u(x), c\}$, where $u(x)$ is a function and c is a constant. We introduce the operations

$$\{u_1, c_1\} \oplus \{u_2, c_2\} = \biggl\{h \frac{d}{dx} \ln \biggl(\exp \biggl(\frac{1}{h} \biggl(\int_0^x u_1(x)\, dx + c_1 \biggr)\biggr) \\ + \exp \biggl(\frac{1}{h} \biggl(\int_0^x u_2(x)\, dx + c_2 \biggr)\biggr)\biggr), h \ln \biggl(\exp \frac{c_1}{h} + \exp \frac{c_2}{h} \biggr)\biggr\},$$

$$\{u_1, c_1\} \odot \lambda = \{u_1, c_1 + \lambda\},$$

$$\{u_1, c_1\} \odot \{u_2, c_2\} = \{u_1 + u_2, c_1 + c_2\},$$

the involution

$$\{u, c\}^* = \{-u, -c\},$$

and the "inner product" given by the bilinear form

$$(\{u_1, c_1\}^*, \{u_2, c_2\}) = \\ = \ln \int \exp \biggl(\frac{1}{h} \biggl(\int_0^x (u_1(x) - u_2(x))\, dx + (c_2 - c_1) \biggr)\biggr)\, dx.$$

The (matrix) resolving operator for the pair of equations

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \frac{h}{2} \frac{\partial^2 u}{\partial x^2},$$

$$\frac{\partial c}{\partial t} = \frac{h}{2} \frac{\partial u}{\partial x} u_x(0, t) - \frac{1}{2} u^2(0, t)$$

is linear in the described space, eqipped with the structure of a module. Note that one of these equations is the Burgers equation itself. We see that the situation, though more complicated, is essentially the same.

In closing, note that formulas (3.2) and (3.15) for the solution of the Hamilton-Jacobi equation have quite a natural interpretation in idempotent analysis. The function $S(t, x, \xi)$ in (3.2), which is a solution of the Cauchy problem with the initial data given by the idempotent δ-function (3.4), is an analog of the Green function in the theory of linear partial differential equations. Equation (3.2), which can be rewritten via the idempotent integral in the form

$$S(t, x) = \int_{\mathbb{R}^n}^{\oplus} S(t, x, \xi) \odot S_0(\xi)\, d\xi,$$

is an analog of the Green function representation of the solution to a linear Cauchy problem, and Eq. (3.15), rewritten in the equivalent form

$$T(x) = \int_{t \le 0}^{\oplus} t \odot S(t, x, \xi)\, d\mu(t),$$

is a linear formula that expresses the stationary solution via the Green function $S(t, x, \xi)$.

3.2. Generalized Solutions of the Hamilton–Jacobi–Bellman Equation and Eigenfunctions of the Resolving Operator of the Cauchy Problem

For simplicity, we first consider equations that do not depend on the argument x explicitly. We begin with constructing a supply of classical (smooth) solutions by the method of generalized (subgradient) characteristics.

Consider the Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H\left(\dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x), \end{cases} \tag{3.26}$$

where $x \in \mathbb{R}^n$ and $H: \mathbb{R}^n \to \mathbb{R}$ is a convex function (we do not require H to be strictly convex). Note that H is continuous, which follows from the fact that it is finite and convex, but no additional smoothness conditions are imposed. Let us also note that in control theory H is usually represented in the form

$$H(p) = \max_{(f,g)\in F} (\langle f, p\rangle + g), \tag{3.27}$$

where F is some compact set in $\mathbb{R}^{n+1}$.

Let $L(v)$ be the Legendre transform of $H(p)$. Then L is also convex, but it can be discontinuous and assume the value $+\infty$. Let us define an operator family $\{R_t\}_{t>0}$ acting on functions $S_0(x)$ bounded below by setting

$$(R_t S_0)(x) = S(t, x) = \inf_{\xi\in\mathbb{R}^n} \left(S_0(\xi) + tL\left(\frac{x-\xi}{t}\right)\right). \tag{3.28}$$

The operators R_t are obviously linear with respect to the operations $\oplus = \min$ and $\odot = +$ and are characterized by the property that their kernels (see §3.1) are convex functions of the difference of the arguments.

Proposition 3.1 ([?]) *Let $S_0(x)$ be smooth and strongly convex, so that for all x the eigenvalues of the Hessian matrix $S_0''(x)$ are not less than some $\delta > 0$. Then*

1) *For any $x \in \mathbb{R}^n$ and $t > 0$ there exists a unique $\xi(t, x) \in \mathbb{R}^n$ such that $(x - \xi(t, x))/t$ is a subgradient of H at the point $S_0'(\xi(t, x))$; moreover,*

$$(R_t S_0)(x) = S_0(\xi(t, x)) + tL\left(\frac{x - \xi(t, x)}{t}\right). \tag{3.29}$$

2) *For $t > 0$ the mapping $\xi(t, x)$ is Lipschitz continuous on compact sets, and $\lim_{t\to 0} \xi(t, x) = x$.*

3) *The Cauchy problem* (3.26) *has a unique* C^1 *solution. This solution is given by Eq.* (3.28) *or Eq.* (3.29) *and satisfies*

$$\frac{\partial S}{\partial x}(t,x) = S_0'(\xi(t,x)).$$

Remark 3.1 If H is a homogeneous function of the form (3.27), then Eq. (3.29) acquires the especially simple form

$$R_t S_0(x) = \min\{S_0(\xi) : \xi \in x - t \operatorname{co} F\},$$

where $\operatorname{co} F$ is the convex hull of a set F.

Remark 3.2 Proposition 3.1 is quite obvious if H is smooth and strictly convex, since in this case L is also smooth and all subgradients involved are usual gradients.

We postpone the proof of Proposition 3.1, which is rather lengthy and technical, until stating Theorem 3.1 so as not to interrupt the exposition.

As was noted in Chapter 2, smooth convex functions form a "complete" subset in $C_{\mathbb{0}}^\infty(\mathbb{R}^{2n})$, since they approximate the "δ-function"

$$\delta_\xi(x) = \lim_{n\to\infty} n(x-\xi)^2 = \begin{cases} \mathbb{1} = 0, & x = \xi, \\ \mathbb{0} = +\infty, & x \neq \xi. \end{cases}$$

Consequently, each functional $\varphi \in (C_{\mathbb{0}}^\infty(\mathbb{R}^n))^*$ is uniquely determined by its values on this set of functions.

The Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H\left(-\dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.30}$$

with Hamiltonian $\widetilde{H}(p) = H(-p)$ will be called the *adjoint problem* of the Cauchy problem (3.26). Obviously, the classical resolving operator R_t^* of the Cauchy problem (3.29), which is determined on smooth convex functions by the formula

$$(R_t^* S_0)(x) = \inf_\xi \left(S_0(\xi) + tL\left(\frac{\xi - x}{t}\right)\right), \tag{3.31}$$

is linear (with respect to the operations $\oplus = \min$ and $\odot = +$) on this set of functions and is the adjoint of the resolving operator R_t of the Cauchy problem (3.26) with respect to the inner product introduced in §1.2. We are now in a position to define weak solutions of problem (3.26) by analogy with the theory of linear equations; we also take into account the fact that, by Theorem 2.1.2, the functionals $\varphi \in (C_{\mathbb{0}}^\infty(\mathbb{R}^n))^*$ are given by usual functions bounded below.

Definition 3.1 Let $S_0 : \mathbb{R}^n \to A = \mathbb{R} \cup \{+\infty\}$ be a function bounded below, and let $m_{S_0} \in (C_{\mathbb{0}}^\infty(\mathbb{R}^n))^*$ be the corresponding functional. The *generalized* *weak solution* of the Cauchy problem (3.26) is the function $(R_tS_0)(x)$ determined by the equation

$$m_{R_tS_0}(\psi) = m_{S_0}(R_t^*\psi)$$

for all smooth strictly convex functions ψ.

The following theorem is a direct consequence of Definition 3.1, Theorem 2.2, and Proposition 3.1.

Theorem 3.1 *For an arbitrary function $S_0(x)$ bounded below, the generalized weak solution of the Cauchy problem* (3.26) *exists and can be found according to the formula*

$$(R_tS_0)(x) = \inf_\xi \left(S_0(\xi) + tL\left(\frac{x-\xi}{t}\right)\right).$$

Various solutions have the same lower semicontinuous closure Cl*, so the solution in the class of semicontinuous functions is unique and is given by the formula*

$$\mathrm{Cl}(R_tS_0) = R_t \, \mathrm{Cl}\, S_0.$$

Proof of Proposition 3.1. 1) Let us prove the uniqueness of a point $\xi(t,x)$ with the desired property. We use the following general inequality, which readily follows from the definition of a subgradient [**?**]: if α_1 and α_2 are subgradients of $F(u)$ at points u_1 and u_2, respectively, then

$$\langle \alpha_1 - \alpha_2, u_1 - u_2 \rangle \geq 0,$$

and the inequality is strict if $F(u)$ is strictly convex.

Suppose that there exist two points $\xi_1 \neq \xi_2$ such that $\alpha_i = (x-\xi_i)/t$ are subgradients of H at $S_0'(\xi_i)$, $i = 1, 2$. Then

$$x = \xi_1 + t\alpha_1 = \xi_2 + t\alpha_2.$$

By the cited inequality,

$$\begin{aligned}&\langle (\xi_1 + t\alpha_1) - (\xi_2 + t\alpha_2), S_0'(\xi_1) - S_0'(\xi_2)\rangle \\ &\qquad = \langle \xi_1 - \xi_2, S_0'(\xi_1) - S_0'(\xi_2)\rangle + t\langle \alpha_1 - \alpha_2, S_0'(\xi_1) - S_0'(\xi_2)\rangle > 0.\end{aligned}$$

This contradiction proves that $\xi(t,x)$ is unique.

To prove the existence of $\xi(t,x)$ and Eq. (3.29), let us show that if $\xi(t,x)$ is the point at which the infimum in Eq. (3.28) is attained (this point exists and is unique since $S_0(x)$ is strongly convex), then $(x - \xi(t,x))/t$ is a subgradient of H at $S_0'(\xi)$.

The point $\xi = \xi(t,x)$ of infimum in Eq. (3.28) can be characterized by the variational inequality [**?**]

$$\langle S_0'(\xi), \eta - \xi\rangle + tL\left(\frac{x-\eta}{t}\right) - tL\left(\frac{x-\xi}{t}\right) \geq 0 \qquad \forall \eta. \tag{3.32}$$

We write

$$\psi(\xi) = \frac{x-\xi}{t}, \qquad \varphi = \frac{x-\eta}{t}$$

and find that inequality (3.32) is equivalent to the inequality

$$\langle S_0'(\xi), \psi(\xi) - \varphi\rangle + L(\varphi) - L(\psi(\xi)) \geq 0,$$

or

$$\langle S_0'(\xi), \psi(\xi)\rangle - L(\psi(\xi)) \geq \langle S_0'(\xi), \varphi\rangle - L(\varphi),$$

for all φ. Since H is the Legendre transform of L, it follows from the last inequality that its left-hand side is equal to $H(S_0'(\xi))$, that is,

$$H(S_0'(\xi)) + L(\psi(\xi)) = \langle S_0'(\xi), \psi(\xi)\rangle. \tag{3.33}$$

The last equation is characteristic of the subgradient [**?**], whence it follows that $\psi(\xi) = (x-\xi)/t$ is a subgradient of H at $S_0'(\xi)$, as desired.

2) By using the variational inequalities (3.32) at the points $\xi = \xi(t,x)$ and $\eta = \xi(\tau, y)$, we obtain

$$\left\langle S_0'(\xi), \frac{x-\xi}{t} - \frac{y-\eta}{\tau}\right\rangle + L\left(\frac{y-\eta}{\tau}\right) - L\left(\frac{x-\xi}{t}\right) \geq 0,$$
$$\left\langle S_0'(\eta), \frac{y-\eta}{\tau} - \frac{x-\xi}{t}\right\rangle + L\left(\frac{x-\xi}{t}\right) - L\left(\frac{y-\eta}{\tau}\right) \geq 0.$$

Adding these inequalities yields

$$\langle S_0'(\xi) - S_0'(\eta), \tau x - ty\rangle \geq \langle S_0'(\xi) - S_0'(\eta), \tau\xi - t\eta\rangle, \tag{3.34}$$

or

$$\langle S_0'(\xi) - S_0'(\eta), \tau(x-y) + (\tau - t)(y-\eta)\rangle \geq \tau\langle S_0'(\xi) - S_0'(\eta), \xi - \eta\rangle. \tag{3.35}$$

It is easy to see that if (x,t) and (y,t) belong to some compact subset P of the domain $\{t > 0\}$, then ξ and η also lie in some compact set (this was proved in the beginning of item 3). Hence, by using the inequalities

$$\langle S_0'(\xi) - S_0'(\eta), \xi - \eta\rangle \geq \delta\|\xi - \eta\|^2,$$
$$\|S_0'(\xi) - S_0'(\eta)\| \leq 2\|\xi - \eta\| \max_{\theta\in[\xi,\eta]} \|S_0''(\theta)\|,$$

from Eq. (3.33) we obtain

$$\delta\|\xi - \eta\|^2\tau \leq 2\lambda(\tau\|x-y\| + (\tau - t)\|y - \eta\|)\|\xi - \eta\|,$$

for $t > \tau$, where λ is some constant depending on P. It follows that $\xi(t,x)$ is Lipschitz continuous on compact subsets in $\{t > 0\}$.

Obviously, $\|\xi(t,x)\|$ is bounded if x is fixed and $t \le t_0$. Hence, should $\xi(t,x) \to y \ne x$ as $t \to 0$ along some sequence, the norm of the subgradient $(x - \xi(t,x))/t$ along the same sequence would tend to infinity, which is impossible since the set of subgradients of $H(p)$ in each compact neighborhood of $S_0'(y)$ is bounded. We conclude that $\xi(t,x) \to x$ as $t \to 0$.

3) a) By (3.33), Eq. (3.29) can be rewritten in the form

$$S(t,x) = S_0(\xi) + \langle S_0'(\xi), x - \xi\rangle - tH(S_0'(\xi)),$$

where $\xi = \xi(t,x)$. It follows, in view of 2), that $S(t,x)$ is continuous and the initial conditions are satisfied.

b) Clearly, $L(v) \to +\infty$ as $\|v\| \to +\infty$, since

$$L(v) \ge \max_{\|p\|=1} (pv - H(p)).$$

Hence, there exists an $A > 0$ such that $L(v) > L(0)$ whenever $\|v\| \ge A$.

Let P be a compact set in the half-space $\{x \in \mathbb{R}^n,\ t > 0\}$. Then there exists a compact set $K \subset \mathbb{R}^n$ such that for $\xi \ne K$ and $(t,x) \in P$ we have $S_0(\xi) > S_0(x)$, $\|(x-\xi)/t\| \ge A$, and hence,

$$S_0(\xi) + tL\left(\frac{x-\xi}{t}\right) > S_0(x) + tL(0).$$

It follows that if $(t,x) \in P$, then $\xi(t,x) \in K$, and consequently,

$$S(t,x) = \min_{\xi \in K}\left(S_0(\xi) + tL\left(\frac{x-\xi}{t}\right)\right).$$

c) Let us now take some $B > A$ such that the ball

$$Q_B = \{v : \|v\| \le B\}$$

contains the set

$$\{v = (x-\xi)/t : (t,x) \in P,\ \xi \in K\}$$

and construct a sequence of smooth strictly convex functions $L_n\colon \mathbb{R}^n \to \mathbb{R}$ so that in Q_B this sequence is monotone increasing and pointwise convergent to L. Such a sequence can be constructed, say, by smoothing (integral averaging of) the sequence

$$\widetilde{L}_n(v) = \max_{\|p\| \le n} (pv - H(p)),$$

which consists of continuous convex everywhere finite functions and is pointwise monotonically convergent to L. By discarding a few first terms where necessary, we can assume that $L_1(v) > L(0)$ for $\|v\| = A$, and hence, $L_n(v) > L_n(0)$ for all n for $\|v\| = A$.

Let H_n be the Legendre transform of L_n. Then H_n is smooth and strictly convex, and so the statement of Lemma 3.1 is trivial for the corresponding Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H_n\left(\dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x). \end{cases} \tag{3.36}$$

Let $\xi_n(t,x)$ denote the mapping associated with H_n by virtue of item 1) of this lemma. Then the formula

$$S_n(t,x) = S_0(\xi_n(t,x)) + L_n\left(\frac{x-\xi_n(t,x)}{t}\right)$$

specifies a smooth solution of the Cauchy problem (3.36), and

$$\frac{\partial S_n}{\partial x}(t,x) = S_0'(\xi_n(t,x)).$$

By virtue of the cited assumptions, $\xi_n(t,x) \in K$ for any n whenever $(t,x) \in P$. Hence,

$$S_n(t,x) = \min_{\xi \in K}\left(S_0(\xi) + tL_n\left(\frac{x-\xi}{t}\right)\right).$$

Thus, for $(t,x) \in P$ the sequence $S_n(t,x)$ of smooth functions is monotone nondecreasing as $n \to \infty$ and is bounded above by $S(t,x)$. We see that this sequence has a limit

$$\widetilde{S}(t,x) = \lim_{n\to\infty} S_n(t,x) \le S(t,x)$$

for $(t,x) \in P$.

d) Let $(t,x) \in P$. Let us prove that

$$\xi_n = \xi_n(t,x) \to \xi = \xi(t,x)$$

as $n \to \infty$ and that $\widetilde{S}(t,x) = S(t,x)$. Consider an arbitrary subsequence of ξ_n convergent to a point $\eta_0 \in K$. The assumption $L((x-\eta_0)/t) = +\infty$ immediately results in a contradiction, and so in what follows we assume that $L((x-\eta_0)/t)$ is finite. Then for any $\varepsilon > 0$ there exists a positive integer N such that

$$L_n\left(\frac{x-\eta}{t}\right) > L\left(\frac{x-\eta}{t}\right) - \varepsilon \tag{3.37}$$

for $n > 0$. Then inequality (3.37) obviously holds for all η in some neighborhood of η_0. Hence,

$$S_n(t,x) = S_0(\xi_n) + tL_n\left(\frac{x-\xi_n}{t}\right) > S_0(\xi_n) + t\left(L\left(\frac{x-\eta_0}{t}\right) - \varepsilon\right)$$

for large n. Since ε is arbitrary, by passing to the limit we obtain

$$\widetilde{S}(t,x) \geq S_0(\eta_0) + tL\left(\frac{x-\eta_0}{t}\right) \geq S(t,x),$$

whence it follows that $\widetilde{S}(t,x) = S(t,x)$ and $\eta_0 = \xi$.

e) Since the sequence S_n is monotone and the functions S_n and S are continuous, it follows that the convergence $S_n \to S$ is uniform on the compact set P. Hence, so is the convergence $\xi_n \to \xi$; indeed,

$$\begin{aligned}
&|S(t,x) - S_n(t,x)| = S(t,x) - S_n(t,x) \\
&\quad = S_0(\xi) + L\left(\frac{x-\xi}{t}\right) - S_0(\xi_n) - L_n\left(\frac{x-\xi_n}{t}\right) \\
&\quad\quad - L_n\left(\frac{x-\xi}{t}\right) + L_n\left(\frac{x-\xi}{t}\right) \\
&\quad \geq \left(S_0(\xi) + L_n\left(\frac{x-\xi}{t}\right)\right) - \left(S_0(\xi_n) + L_n\left(\frac{x-\xi_n}{t}\right)\right) \geq \frac{\delta}{2}\,\|\xi - \xi_n\|^2.
\end{aligned}$$

f) Thus, the smooth functions $S_n(t,x)$ are uniformly convergent to $S(t,x)$ on P, and their derivatives

$$\frac{\partial S_n}{\partial x}(t,x) = S_0'(\xi_n(t,x))$$

and

$$\begin{aligned}
\frac{\partial S_n}{\partial t} &= -H_n(S_0'(\xi_n(t,x))) = L_n\left(\frac{x-\xi_n}{t}\right) - \left\langle S_0'(\xi_n), \frac{x-\xi_n}{t}\right\rangle \\
&= \frac{1}{t}\left[S_n(t,x) - S_0(\xi_n) - \langle S_0'(\xi_n), x-\xi_n\rangle\right]
\end{aligned}$$

are uniformly convergent to the functions $S_0'(\xi(t,x))$ and $-H(S_0'(\xi(t,x)))$, respectively, by virtue of e). It follows that the latter functions are the derivatives of $S(t,x)$, and so $S(t,x)$ is a C^1 solution of the Cauchy problem (3.26).

g) The uniqueness of a smooth solution is a standard consequence of the fact that $H(p)$ is locally Lipschitz continuous.

Proposition 3.1 is proved.

Let us now proceed to the Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H\left(x, \dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.38}$$

with general Hamiltonian H convex in p. We define generalized solutions for the case of a smooth function H by analogy with problem (3.26); for nonsmooth H, a limit procedure will be applied.

Let H satisfy the following conditions.

1) H is C^2.

2) The second derivatives of H are uniformly bounded:

$$\max\left(\sup_{x,p}\left\|\frac{\partial^2 H}{\partial x^2}\right\|, \sup_{x,p}\left\|\frac{\partial^2 H}{\partial x \partial p}\right\|, \sup_{x,p}\left\|\frac{\partial^2 H}{\partial p^2}\right\|\right) \leq c = \text{const}\,.$$

3) H is strongly convex; that is, there exists a constant $\delta > 0$ such that the least eigenvalue of the matrix $\partial^2 H/\partial p^2$ is greater than or equal to δ for all (x, p).

By $(y(\tau, \xi, p_0), p(\tau, \xi, p_0))$ we denote the solution of Hamilton's system

$$\begin{cases} \dot{y} = \dfrac{\partial H}{\partial p}, \\ \dot{p} = -\dfrac{\partial H}{\partial x} \end{cases} \tag{3.39}$$

with the initial conditions $y(0) = \xi$, $p(0) = p_0$. It follows from theory of ordinary differential equations that under the stated conditions the solution exists and is unique for all $t \geq 0$ and for any (ξ, p_0). Let $S(t, x, \xi)$ denote the greatest lower bound of the action functional

$$\int_0^t L(z(\tau), \dot{z}(\tau))\, d\tau \tag{3.40}$$

over all continuous piecewise smooth curves joining ξ with x in time t ($z(0) = \xi$ and $z(t) = x$). Here $L(x, v)$ is the Lagrangian of the variational problem associated with H, that is, the Legendre transform of H with respect to the variable p.

The following assertion comprises some useful facts from calculus of variations.

Proposition 3.2 *Let H satisfy the above-stated conditions. Then*

a) *The derivatives of the solution of system* (3.36) *with respect to the initial data are given by the formulas*

$$\begin{cases} \dfrac{\partial y}{\partial p_0} = \displaystyle\int_0^t \frac{\partial^2 H}{\partial p^2}(y(\tau), p(\tau))\, d\tau + O(t^2), \\ \dfrac{\partial p}{\partial p_0} = E + O(t), \end{cases} \tag{3.41}$$

where E is the identity matrix and the estimates $O(t)$ and $O(t^2)$ are uniform with respect to ξ, p_0, and $t \in [0, t_0]$ for each fixed t_0.

b) *There exists a $t_0 > 0$ such that for any $t \leq t_0$, ξ, and x the equation $y(t, \xi, p_0) = x$ has a unique solution $p_0 = p_0(t, x, \xi)$. This solution provides* *minimum in the corresponding variational problem with fixed endpoints, that is,*

$$S(t,x,\xi)=\int_0^t L(y(\tau),\dot y(\tau))\,d\tau,$$

where $y(\tau)=y(\tau,\xi,p_0(t,x,\xi))$.

c) *For* $t\le t_0$, *the function* $S(t,x,\xi)$ *defined in* b) *is twice continuously differentiable and strictly convex both in* ξ *and in* x.

Scheme of the proof. Equations (3.41) can be obtained by analyzing the variational system along the solutions of system (3.39). Let us prove b). It follows from Eq. (3.41) that

$$\frac{\partial y}{\partial p_0}=tY(t)+O(t^2),$$

where the matrix $Y(t)$ is a smooth function of t, p_0, and ξ for $t>0$, is symmetric, and has eigenvalues that all lie in the interval $[\delta,c]$. It follows that the matrix $\partial y/\partial p_0$ is invertible for $t\le t_0$ for some fixed t_0. Then, by the inverse function theorem, for any fixed $t\le t_0$ and $\xi\in\mathbb{R}^n$ the mapping

$$P\colon\mathbb{R}^n\to\mathbb{R}^n,\qquad p_0\mapsto y(t,\xi,p_0)$$

is locally invertible in a neighborhood of each $p_0\in\mathbb{R}^n$; moreover, $P(p_0)\to\infty$ as $p_0\to\infty$. We see that the range of P is open and closed in $\mathbb{R}^n$ and hence coincides with $\mathbb{R}^n$, so that $P\colon\mathbb{R}^n\to\mathbb{R}^n$ is surjective. Consequently, P is a covering and, what is more, the universal covering, since $\mathbb{R}^n$ is simply connected. It follows from the properties of universal coverings that P is a bijection. Indeed, the preimage of each point (the fiber of the universal covering) is an orbit of the monodromy group, which is isomorphic to the fundamental group of the space $\mathbb{R}^n$ (the base of the covering) and hence trivial. Thus, we have proved that for any x and ξ there exists a unique p_0 such that $y(t,\xi,p_0)=x$. The fact that this solution provides a minimum of the action functional follows from the standard criteria in calculus of variations. Indeed, the extremal $y(\tau)$ can be included in a field of extremals and the Lagrangian L is convex in the second argument. Item c) now follows from the well-known formula

$$\frac{\partial^2 S}{\partial x^2}(t,x,\xi)=\frac{\partial p(t,\xi,p_0(t,x,x))}{\partial x}=\frac{\partial p}{\partial p_0}\left(\frac{\partial y}{\partial p_0}\right)^{-1} \tag{3.42}$$

for the derivatives of S. By Proposition 3.2, if the initial function $S_0(x)$ in the Cauchy problem (3.38) is convex, then the function

$$(R_tS_0)(x)=S(t,x)=\min_\xi(S_0(\xi)+S(t,x,\xi)) \tag{3.43}$$

is continuously differentiable for all $t \le t_0$ and $x \in \mathbb{R}^n$. Indeed, the minimum in (3.43) is obviously attained at a unique point $\xi(t,x)$. Therefore, Eq. (3.43) specifies the unique classical solution of the Cauchy problem (3.38). Thus, the technique of determining the usual solutions of the Cauchy problem (3.38) for $t \le t_0$ for an arbitrary initial function $S_0(x)$ via the resolving operator of the adjoint Cauchy problem

$$\begin{cases} \dfrac{\partial S}{\partial t} + H\left(x, -\dfrac{\partial S}{\partial x}\right) = 0, \\ S|_{t=0} = S_0(x) \end{cases} \tag{3.44}$$

is just a reproduction of the method presented above for Hamiltonians independent of x. For $t \ge t_0$, the resolving operator (3.38) is defined via iterations of the resolving operator given for $t \le t_0$. As a result, we conclude that the following assertion is valid.

Theorem 3.2 *Suppose that the Hamiltonian H satisfies the above-stated conditions. Then for any function $S(x)$ bounded below the Cauchy problem* (3.35) *is solvable and the solution is unique modulo functions with the same lower semicontinuous closure. The solution is equal to $R_t \operatorname{Cl} S_0 = \operatorname{Cl} R_t S_0$, where*

$$(R_t S_0)(x) = \inf_{\xi}(S_0(\xi) + S(t,x,\xi)). \tag{3.45}$$

Furthermore, it is obvious that if a smooth function of the form $S(t,x) = S_0(x) + t\lambda$ is a solution of problem (3.38), then $S_0(x)$ is a classical solution of the stationary equation

$$H\left(x, \frac{\partial S}{\partial x}\right) + \lambda = 0. \tag{3.46}$$

Thus, it is natural to define a generalized solution of Eq. (3.46) as an eigenfunction (in the sense of idempotent analysis) of the resolving operator (3.45) of the nonstationary problem. Let the Lagrangian $L(x,v)$, defined as the Legendre transform of $H(x,p)$ with respect to p, satisfy $L(x,v) \to \infty$ as $\|x\|, \|v\| \to \infty$. Then the operator R_t (3.45) is a compact A-linear operator and has at most one eigenvalue by Theorem 2.3.2. It turns out that in this case there is a natural method of constructing eigenfunctions of R_t (generalized solutions of Eq. (3.46)). We consider the important particular case in which the semimodule of generalized solutions is finite-dimensional. This is just the situation occurring when multiplicative asymptotics are studied in Chapter 4.

Theorem 3.3 ([?]) *Suppose that $L(x,v)$ has finitely many points of minimum $(\xi_1, 0), \dots, (\xi_k, 0)$. Then $\lambda = \min_{x,v} L(x,v)$ is a value for which there exist generalized solutions of the stationary problem* (3.46); *the semimodule* *of these solutions (in the sense of idempotent structure) has a finite basis* $\{S_1, \dots, S_k\}$*, where* $S_j(x)$ *is the infimum of the functional*

$$J_j(q(\,\cdot\,)) = \int_0^t (L(q(\tau), \dot{q}(\tau)) - \lambda)\, dt \tag{3.47}$$

over all $t > 0$ *and all piecewise smooth curves joining* ξ_j *with* x *in time* t*. Moreover, the operator family* $R_t - \lambda t$ *converges as* $t \to 0$ *to the finite-dimensional operator* B *given by the formula*

$$(Bh)(x) = \bigoplus_j \langle h, \widetilde{S}_j \rangle_A \odot S_j(x), \tag{3.48}$$

where $\{\widetilde{S}_1, \dots, \widetilde{S}_k\}$ *is a basis of the eigensemimodule for the adjoint operator.*

Proof. Let us show that each S_j is an eigenfunction of the resolving operator (3.18):

$$\begin{aligned}(R_t S_j)(x) &= \inf_\xi \inf_{\tau \geq 0} (S(\tau, \xi, \xi_j) - \tau\lambda + S(t, x, \xi)) \\ &= \inf_{\tau \geq 0} (S(t + \tau, x, \xi_j) - \tau\lambda) = \inf_{\tau \geq t} (S(\tau, x, \xi_j) - \tau\lambda + t\lambda) \\ &= \inf_{\tau \geq 0} (S(\tau, x, \xi_j) + \tau\lambda) + t\lambda = S_j(x) + \lambda t.\end{aligned}$$

The limit equation (3.48) means that the family $R_t - \lambda t$ is convergent to the operator with the decomposable kernel

$$\bigoplus_{j=1}^k S_j(x) \odot \widetilde{S}_j(y),$$

which can be proved by analogy with Theorem 2.3.4.

Let us return to the general Cauchy problem (3.35). Suppose that the Hamiltonian H is representable as the limit (uniform on compact sets) of a sequence of Hamiltonians H_n satisfying the assumptions of Theorem 3.2. It is clear that if we define generalized solutions of problem (3.35) as the limits of the corresponding solutions of the Cauchy problems with Hamiltonians H_n, then Theorem 3.2 will be valid for H. In optimal control theory, the Hamiltonians $H(x, p)$ homogeneous in p, namely, those given by the formula

$$H(x, p) = \max_{u \in U} \langle f(x, u), p \rangle, \tag{3.49}$$

are of special importance, where U is a compact set and $f(x, u)$ is a function continuous on $\mathbb{R}^n \times U$ and Lipschitz continuous in x uniformly with respect to $u \in U$:

$$|f(x_1, u) - f(x_2, u)| \leq L\|x_1 - x_2\|$$

for any x_1, $x_2 \in \mathbb{R}^n$, any $u \in U$, and some L. As was indicated in §3.1, problems with Hamiltonians of this sort arise when dynamic programming is used to find optimal trajectories of controlled objects whose motion is described by the law $\dot{x} = f(x, u)$. For homogeneous Hamiltonians, Eq. (3.43) can be rewritten in the special form

$$(R_t S_0)(x) = \inf\{S_0(\xi) : \xi \in K(t, x)\}, \tag{3.50}$$

where $K(t, x)$ is the t-slice of the integral funnel of the generalized differential equation $\dot{y} \in -F(y)$, issuing from x at $t = 0$. Here $F(y)$ is the convex hull of the compact set $\{f(y, u) : u \in U\}$. This formula follows from Eq. (3.43) and from the fact that the Legendre transform of the function H (3.49) has the form

$$L(x, v) = \begin{cases} 0, & v \in F(x), \\ +\infty, & v \notin F(x), \end{cases}$$

so that the action functional (3.38) in this situation is equal to zero on the solutions of the generalized differential equation $\dot{x} \in F(x)$ and is infinite on any other trajectory. Equation (3.50) can also be obtained directly with the aid of the adjoint operator, as in Theorem 3.1 (see [**?**]).

The differential Bellman equation is formally obtained by passing to the limit in the discrete Bellman equation (cf. §3.1). We can ascribe exact meaning to this passage to the limit by using the cited notion of a generalized solution and linearity.

Let us define an operator Δ_τ on functions $\varphi(x)$ bounded below by the formula

$$(\Delta_\tau \varphi)(x) = \inf_u \varphi(x - f(x, u)\tau).$$

For each initial function $S_0(x)$ of this class, we consider the sequence of functions

$$S_k(t, x) = (\Delta_\tau^k S_0)(x), \qquad \tau = t/k. \tag{3.51}$$

Lemma 3.1 *If S_0 is upper semicontinuous and bounded below, then the sequence S_k is pointwise convergent to the function* (3.50)*, which is also upper semicontinuous.*

The proof follows from the possibility to approximate solutions of generalized differential equations by Euler's polygons and from the definition of semicontinuous functions.

Theorem 3.4 *If $S_0(x)$ is bounded below, then the sequence $S_k(t, x)$ of the form* (3.51) *is weakly convergent in the sense of the inner product* (*introduced in Chapter* 1) *on the set of functions upper semicontinuous and bounded below to the weak solution $S(t, x)$ of the Cauchy problem* (3.38) *with the Hamiltonian* (3.49).

Proof. By applying Lemma 3.1 to the adjoint operator

$$(\Delta_\tau^*\varphi)(x) = \inf_u \varphi(x + f(x,u)\tau),$$

which arises when the adjoint equation (3.44) is discretized, for any upper semicontinuous function φ bounded below we obtain

$$\begin{aligned}\langle S_k(t,x), \varphi\rangle_A &= \langle \Delta_\tau^k S_0, \varphi\rangle_A \\ &= \langle S_0, (\Delta_\tau^*)^k\varphi\rangle_A \overset{k\to+\infty}{\longrightarrow} \langle S_0, R_t^*\varphi\rangle_A = \langle R_t S_0, \varphi\rangle_A,\end{aligned}$$

as desired.

In conjunction with Theorem 1.4.3, this theorem implies the following.

Corollary 3.1 a) *The lower semicontinuous solution $R_t(\operatorname{Cl} S_0) = S(t,x)$ of the Cauchy problem* (3.38) *with homogeneous Hamiltonian* (3.51) *is the lower envelope of the sequence $S_k(t,x)$.*

b) *For each open $\Omega \subset \mathbb{R}^n$ one has*

$$\lim_{k\to\infty} \inf_{x\in\Omega} S_k(t,x) = \inf_{x\in\Omega} S(t,x).$$

If $S_0(x)$ is continuous, then $S_k(t,x)$ converges to $S(t,x)$ in the usual sense. Let us state and prove this result for completeness.

Proposition 3.3 *If $S_0(x)$ is a Lipschitz continuous function with Lipschitz constant L_0, then all functions $S_k(t,x)$ are continuous and Lipschitz continuous in x with Lipschitz constant $L_0 e^{Lt}$, where the constant L is introduced after Eq.* (3.51); *the sequence $S_k(t,x)$ is convergent to the function* (3.50) *uniformly on compact sets.*

For simplicity, we carry out the proof for binary partitions, i.e., for the subsequence S_{k_l}, where $k_l = 2^l$.

1) Obviously,

$$S_k(t,x) = \min_{u_1,\dots,u_k} S_0(x_k(\tau, x; u_1, \dots, u_k)),$$

where $k = 2^l$ and the x_m are specified by the recursion equations

$$x_m = x_{m-1} - f(x_{m-1}, u_m)\tau, \qquad \tau = t/k,\ x_0 = x.$$

Furthermore,

$$\|f(x,u)\| \le M + L\|x\|,$$

whence it follows by induction that the estimate

$$\|x_k(\tau, x; u_1, \dots, u_k) - x\| \le \frac{M + L\|x\|}{L}\,[(1 + L\tau)^k - 1]$$

is valid, and consequently,

$$\|x_k(\tau, x; u_1, \dots, u_k) - x\| \le \frac{M + L\|x\|}{L}\,(e^{Lt} - 1).$$

In particular,

$$\|f(x_k, u)\| \le (M + L\|x\|)e^{Lt}.$$

Let us now estimate the deviation of x_k when the time interval is subdivided. If

$$x'_\tau = x' - f(x', u)\tau,$$
$$x''_\tau = x'' - f(x'', u)\,\frac{\tau}{2} - f\left(x'' - f(x'', u)\,\frac{\tau}{2}, u\right)\frac{\tau}{2},$$

then

$$\begin{aligned}\|x'_\tau - x''_\tau\| &\le \|x' - x''\| + \frac{\tau}{2}\,L\|x' - x''\| + \frac{\tau}{2}\,L\left\|x' - x'' + f(x'', u)\,\frac{\tau}{2}\right\| \\ &\le \|x' - x''\|(1 + L\tau) + \frac{\tau^2}{4}\,L\|f(x'', u)\|.\end{aligned}$$

Consequently,

$$\begin{aligned}&\|x_{2k}(\tau, x; u_1, u_1, \dots, u_k, u_k) - x_k(\tau, x; u_1, \dots, u_k)\| \\ &\qquad \le (1 + L\tau)\|x_{2k-2}(\tau, x; u_1, u_1, \dots, u_{k-1}, u_{k-1}) \\ &\qquad\qquad - x_{k-1}(\tau, x; u_1, \dots, u_{k-1})\| + L(M + L\|x\|)\,\frac{\tau^2}{4}\,e^{Lt} \\ &\qquad \le \cdots \\ &\qquad \le L(M + L\|x\|)\,\frac{\tau^2}{4}\,e^{Lt}(1 + (1 + L\tau) + \dots + (1 + L\tau)^{k-1}) \\ &\qquad = \frac{\tau}{4}\,(M + L\|x\|)e^{Lt}((1 + L\tau)^k - 1) \\ &\qquad \le \frac{\tau}{4}\,(M + L\|x\|)e^{Lt}(e^{Lt} - 1).\end{aligned}$$

It follows from this estimate and from the Lipschitz continuity of $S_0(x)$ that

$$S_{2k}(t, x) \le S_k(t, x) + \frac{L_0}{4k}\,(M + L\|x\|)e^{Lt}(e^{Lt} - 1), \tag{3.52}$$

and hence, the limit $\lim_{l\to\infty} S_{2^l}(t, x)$ exists. Obviously, this limit is equal to $S(t, x)$, since the solutions of generalized differential equations can be approximated by Euler's polygons [**?**].

2) If

$$x'_\tau = x' - f(x', u)\tau', \qquad x''_\tau = x'' - f(x'', u)\tau'',$$

then

$$\|x'_\tau - x''_\tau\| \le (1 + L\tau')\|x' - x''\| + |\tau' - \tau''| \cdot \|f(x'', u)\|.$$

Thus,

$$\begin{aligned}
&\|x_k(t, x; u_1, \dots, u_k) - x_k(t', x'; u_1, \dots, u_k)\| \\
&\quad \le (1 + L\tau')\|x_{k-1}(t, x; u_1, \dots, u_{k-1}) - x_{k-1}(t', x'; u_1, \dots, u_{k-1})\| \\
&\qquad + |\tau' - \tau''|(M + L\|x\|)e^{Lt} \\
&\quad \le \cdots \\
&\quad \le (1 + L\tau')^k \|x - x'\| \\
&\qquad + \|\tau' - \tau''|(M + L\|x\|)e^{Lt}(1 + (1 + L\tau') + \cdots + (1 + L\tau')^{k-1}) \\
&\quad \le e^{L\tau'}\Delta x + (M + L\|x\|)e^{Lt}\Delta t \, \frac{e^{Lt'} - 1}{Lt'},
\end{aligned}$$

where

$$\Delta t = |t - t'|, \qquad \Delta x = \|x - x'\|, \qquad \tau' = t'/k.$$

It follows that

$$\|S_k(t, x) - S_k(t', x')\| \le L_0 \left(e^{Lt'}\Delta x + (M + L\|x\|)e^{Lt}\Delta x \, \frac{e^{Lt'} - 1}{Lt'} \right),$$

and hence, the functions $S_k(t, x)$ are continuous and Lipschitz continuous in x. In turn, in conjunction with the almost monotone decrease (3.52), this implies that the convergence $S_n \to S$ is uniform.

3.3. Jump Stochastic Perturbations of Deterministic Optimization Problems. An Applications to a Spin Control Problem (Control of a Two-Level Atom)

Idempotent analysis studies resulted in including a series of important nonlinear differential equations (such as numerous optimal control equations and some quasilinear systems occurring in hydrodynamics) in the scope of linear methods, since these equations become linear in the new arithmetic. Idempotent analysis also implies a new approach to the study of a class of nonlinear (even in the new sense) equations, namely, equations "close" to equations linear in idempotent semimodules or semirings. It is natural to study such equations in the framework of the corresponding perturbation theory. Indeed, the theory of numerous important equations of mathematical physics was constructed on the basis of the fact that the nonlinearity is a small "correction" to a linear equation.

The main characteristics of a long-time optimal process are determined by the solutions (λ, h) (where λ is a number and h a function on the state space) of the equation

$$Bh = \lambda + h, \tag{3.53}$$

where B is the Bellman operator of the optimization problem. Namely, λ is the mean income per step of the process, whereas h specifies stationary optimal strategies or even turnpike control modes (see §2.4). For a usual deterministic control problem, in which B is linear in the sense of the operations $\oplus = \min$ or $\oplus = \max$ and $\odot = +$, Eq. (3.53) is the idempotent analog of an eigenvector equation in standard linear algebra. Thus, the solutions λ and h are actually called an *eigenvalue* and an *eigenvector* of B, respectively. In the case of stochastic control, the Bellman operator is no longer linear in the idempotent semimodule in general; it is only homogeneous in the sense of the operation $\odot = +$. However, if the influence exerted on the process by stochastic factors is small (the process is nearly deterministic), then the corresponding Bellman operator is close to an operator linear in the idempotent semimodule, and the solutions to Eq. (3.53) are naturally provided by perturbation theory. The main difficulties encountered in this approach, when we attempt to expand the solution into a series in powers of the small parameter, are due to the fact that B is not smooth. Note that the cited weak stochastic perturbations naturally arise in some models of control over quantum systems and in mathematical economics (for example, when possible random disturbances in the planned production process are taken into account in the model).

Following [**?**], in this section we consider some series of perturbations of Eq. (3.53), linear in the semiring with the operations $\oplus = \max$ and $\odot = +$, when the perturbation theory can be constructed. As a result, we obtain approximate formulas for the mean income and for the stationary optimal strategies in the corresponding controlled stochastic jump processes in discrete or continuous time. As an example, a model of control over a two-level atom (the simplest model of optical pumping in lasers) is considered, where the control is based on registering photons by which the atom interacts with a boson reservoir that models the instrument. The study of this example heavily uses the quantum stochastic filtering theory recently constructed by V. P. Belavkin [**?**]. A similar result for continuous diffusion-type observations was obtained in [**?**].

3.3.1. *Stochastic perturbations of some graph optimization problems*

Let (X, E) be a finite strongly connected graph, where X is the set of nodes, $|X| = n$, and $E \subset X \times X$ is the set of arcs, and let $b\colon E \to \mathbb{R}$ be a weight function. Let a subset $J \subset X$ and a collection $\{q_j^i\}_{i\in X, j\in J}$ of nonnegative numbers be given. In $\mathbb{R}^n$, let us consider the operator family $\{B_\varepsilon\}_{\varepsilon\geq 0}$ given by the formula

$$(B_\varepsilon h)_i = \max_{k:(i,k)\in E}(b(i,k)+h_k)\left(1-\varepsilon\sum_{j\in J}q_j^i\right)+\varepsilon\sum_{j\in J}q_j^i(b(i,j)+h_j).$$

The operator B_ε is the Bellman operator corresponding to a controlled Markov process on (X,E); in this process, the transition from $i\in X$ to $j\in J$ has probability εq_j^i, and the transition to a chosen point k from the set $\{k:(i,k)\in E\}$ occurs with probability $1-\varepsilon\sum_{j\in J}q_j^i$.

Theorem 3.5 *Let the maximum of b be attained at a unique point $(V,V)\in E$, where $b(V,V)=0$, and suppose that $V\notin J$ and there exists a $j\in J$ such that the probability q_j^i is nonzero for all i. Then for each sufficiently small $\varepsilon>0$ there exists a unique solution $\lambda_\varepsilon\in\mathbb{R}$, $h^\varepsilon\in\mathbb{R}^n$ of the equation*

$$B_\varepsilon h^\varepsilon = \lambda_\varepsilon + h^\varepsilon, \qquad h_V^\varepsilon = 0, \tag{3.54}$$

and λ_ε and h^ε are differentiable with respect to ε at $\varepsilon=0$, that is,

$$\lambda_\varepsilon = \varepsilon\lambda' + o(\varepsilon), \qquad h^\varepsilon = h^0+\varepsilon h' + o(\varepsilon).$$

Furthermore, the derivatives λ' and h' are given by the following formulas of perturbation theory:

$$\lambda' = \sum_{j\in J} q_j^V(b(V,j)+h_j^0), \tag{3.55}$$

$$h^i = (\mathrm{Id}-B_h')^{-1}\left(\left\{\sum_{j\in J}q_j^i(b(i,j)+h_j^0-h_i^0)\right\}\right), \tag{3.56}$$

where Id *is the identity operator and*

$$(B_h' g)_i = \max\{g_k : k\in\Gamma_h(i)\}; \tag{3.57}$$

here $\Gamma_h(i)$ is the set of vertices where $b_{ij}+h_j^0$ attains the maximal value.

Proof. It follows from Theorem 2.3.3 that for some n the operator $(B_0)^n$ becomes an operator with separated kernel. Since the operators $(B_\varepsilon)^n$ are close to $(B_0)^n$, it follows that the fixed points of the operators $(\widetilde{B}_\varepsilon)^n$ factorized by constants are close to those of $(\widetilde{B}_0)^n$. Consequently, for each sufficiently small $\varepsilon\ge 0$ there exists a unique solution h^ε of Eq. (3.54) (the existence and uniqueness for $\varepsilon>0$ also follows from the theorems proved in §2.5), and

$$h^\varepsilon = h + \Delta^\varepsilon,$$

where $h=h^0$, $\Delta^\varepsilon = o(1)$, and $\lambda_\varepsilon = o(1)$ as $\varepsilon\to 0$.

Now note that for sufficiently small Δ we have

$$B(h+\Delta)_i = \max_{j\in\Gamma_h(i)} (b(i,j)+h_j+\Delta_j),$$

where $\Gamma_h(i)$ is the set of vertices j at which $b_{ij}+h_j$ attains the maximal value. It follows that

$$B(h+\Delta)_i = Bh_i + (B'_h\Delta)_i, \tag{3.58}$$

where B'_h is defined by formula (3.57).

Thus, Eq. (3.54) can be rewritten in the form

$$B(h+\Delta^\varepsilon)_i + \varepsilon\Bigg[\sum_{j\in J} q^i_j(b(i,j)+h_j+\Delta^\varepsilon_j) - \sum_{j\in J} q^i_j B(h+\Delta^\varepsilon)_i\Bigg] \\ = \lambda_\varepsilon + \varphi_i + \Delta^\varepsilon_i,$$

or, in view of Eq. (3.58) and Eq. (3.54) at $\varepsilon = 0$, in the form

$$(B'_h\Delta^\varepsilon)_i + \varepsilon\Bigg[\sum_{j\in J} q^i_j(b(i,j)+h_j) - \sum_{j\in J} q^i_j(Bh)_i\Bigg] + o(\varepsilon) = \lambda_\varepsilon + \Delta^\varepsilon_i.$$

In other words,

$$(\mathrm{Id} - B'_h)(\Delta^\varepsilon)_i + \lambda_\varepsilon = \varepsilon\Bigg[\sum_{j\in J} q^i_j(b(i,j)+h_j) - \sum_{j\in J} q^i_j(Bh)_i\Bigg] + o(\varepsilon). \tag{3.59}$$

Since $\Gamma_h(V) = V$, we see that the substitution $i = v$ yields

$$\lambda_\varepsilon = \varepsilon\sum_{j\in J} q^v_j(b(v,j)+h_j) + o(\varepsilon).$$

On passing to the factorized operators $\widetilde{B}'_h$ and $\widetilde{B}$ in Eq. (3.59) (see §2.5), we obtain

$$(\mathrm{Id} - \widetilde{B}'_h)\,\frac{\Delta^\varepsilon}{\varepsilon} = \Bigg[\sum_{j\in J} q^i_j(b(i,j)+h_j) - \sum_{j\in J} q^i_j(Bh)_i\Bigg] + o(1).$$

To prove that the limit $\lim_{\varepsilon\to 0}\Delta\varepsilon/\varepsilon$ exists and satisfies Eq. (3.56), it remains to note that the factorized operator $(\mathrm{Id} - B'_h)$ is invertible and that the norm of the inverse does not exceed $|X|$, that is,

$$\|(\mathrm{Id} - \widetilde{B}'_h)^{-1}H\| \le |X|\cdot\|H\|.$$

3.3.2. *Jump stochastic perturbations of dynamic optimization problems on compact sets*

Let the process state space X and the control set U be compact, and let v and V be two distinct points of X. Suppose that the process dynamics is determined by a continuous mapping $y: X \times U \times [0, \varepsilon_0] \to X$ and a continuous function $q: X \to \mathbb{R}_+$ as follows. If a control $u \in U$ is chosen when the process is in a state $x \in X$, then at the current step the transition into the state $y(x, u, \varepsilon)$ takes place with probability $1 - \varepsilon q(x)$, whereas with probability $\varepsilon q(x)$ the transition is into v. The income from residing in a state $x \in X$ is specified by a Lipschitz continuous function $b: X \to \mathbb{R}$. The Bellman operator B_ε obviously acts in the space of continuous functions on X according to the formula

$$(B_\varepsilon h)(x) = b(x) + (1 - \varepsilon q(x)) \max_{u \in U} h(y(x, u, \varepsilon)) + \varepsilon q(x) h(v).$$

Theorem 3.6 *Suppose that for each ε the deterministic dynamics is controllable in the sense that by moving successively from x to $y(x, u, \varepsilon)$ one can reach any point from any other point in a fixed number of steps. Suppose also that b attains its maximum at a unique point V, where $b(V) = 0$ and moreover,*

$$V \in \{y(V, u, \varepsilon) : u \in U\}.$$

Then Eq. (3.53) *is solvable, and the solution satisfies*

$$h^\varepsilon - h^0 = O(\varepsilon), \tag{3.60}$$

$$\lambda_\varepsilon = q(V) h^0(v) \varepsilon + o(\varepsilon), \tag{3.61}$$

where $\lambda_0 = 0$ and h^0 is the unique solution of Eq. (3.53) *at $\varepsilon = 0$.*

The proof is a straightforward generalization of that of Theorem 3.5.

3.3.3. *Generalized solutions of the differential Bellman equation for jump controlled processes on a manifold*

Let X be a smooth compact manifold, and let $f(x, u, \varepsilon)$ be a vector field on X depending on the parameters $u \in U$ and $\varepsilon \in [0, \varepsilon_0]$ and Lipschitz continuous with respect to all arguments. Consider a special case of the process described in subsection 3.3.2, in which $y(x, u, \varepsilon)$ is the point reached at time τ by the trajectory of the differential equation $\dot{z} = f(z, u, \varepsilon)$ issuing from x and the probability of the transition into v in one step of the process is equal to $\tau \varepsilon q(x)$. As $\tau \to 0$, this process becomes a jump process in continuous time; this process is described by a stochastic differential equation with stochastic differential of Poisson type.

Let $S_n^\varepsilon(t, x)$ be the mathematical expectation of the maximal income per n steps of the cited discrete process with time increment $\tau = (T - t)/n$ beginning at time t at a point x and with terminal income specified by a Lipschitz continuous function $S_T(x)$. Then $S_n^\varepsilon = (B_\varepsilon^\tau)^n$, where B_ε^τ is the Bellman operator corresponding to the discrete problem with step τ.

Theorem 3.7 *The sequence of continuous functions S_n^ε is uniformly convergent with respect to x and ε to a Lipschitz continuous (and hence, almost everywhere smooth) function $S^\varepsilon(t,x)$, which satisfies the functional-differential Bellman equation*

$$\frac{\partial S}{\partial t} + b(x) + \varepsilon q(x)(S(v) - S(x)) + \max_{u \in U} \left(\frac{\partial S}{\partial x}, f(x,u,\varepsilon) \right) = 0 \tag{3.62}$$

at each point of differentiability.

The limit function S^ε may be called a generalized solution of the Cauchy problem for Eq. (3.62). This function specifies the mathematical expectation of the optimal income for the limit (as $t \to 0$) jump process in continuous time. For $\varepsilon = 0$, this solution coincides with that obtained in the framework of idempotent analysis (see §3.2). The proof of this theorem, quite lengthy and technical, closely resembles the proof of Proposition 3.3.

Theorems 3.6 and 3.7 imply the following result.

Theorem 3.8 *There exists a continuous function R^ε and a unique λ_ε such that the generalized solution of the Cauchy problem for Eq.* (3.62) *with terminal function $S_T^\varepsilon = h^\varepsilon$ has the form*

$$S^\varepsilon(t,x) = \lambda_\varepsilon(T-t) + h^\varepsilon(x), \tag{3.63}$$

χ_ε satisfies the asymptotic formula (3.61)*, and the generalized solution $S^\varepsilon(t,x)$ of Eq.* (3.62) *with an arbitrary Lipschitz continuous terminal function S_T^ε satisfies the limit equation*

$$\lim_{t \to -\infty} \frac{1}{T-t} S^\varepsilon(t,x) = \lambda_\varepsilon. \tag{3.64}$$

3.3.4. *Example*

As was indicated in the beginning of the section, our results can be useful in approximately solving the Bellman equation corresponding to the controlled dynamics described by a stochastic differential equation. We consider an example of dynamics described by a stochastic equation of Poisson type. An example corresponding to a stochastic equation of diffusion type can be found, e.g., in [**?**].

Consider a model of continuously observed quantum system interacting with an instrument (boson reservoir) by exchanging photons. The a posteriori dynamics (i.e., dynamics taking into account the measurement results) of this system can be described by Belavkin's quantum filtering equation [**?**]

$$d\Phi + \left(i[E,\Phi] + \varepsilon\left(\frac{1}{2}\left(R^*R\Phi + \Phi R^*R\right) - (\Phi, R^*R)\Phi \right)\right) dt = \left(\frac{R\Phi R^*}{(\Phi, R^*R)} - \Phi \right) dN. \tag{3.65}$$

Here $N(t)$ is a counting Poisson process; its spectrum is the set of positive integers, and the result of measurement by time t is a random tuple $\tau = \{t_1 < \cdots < t_n\}$ of time moments at which the photon emission occurs. Furthermore, E and R are closed operators in the Hilbert state space H of the quantum system in question; the self-adjoint energy operator E specifies the free (nonobserved) dynamics, whereas R corresponds to the observed (measured) physical variable. The unknown density matrix Φ specifies the a posteriori (depending on the measurement result τ) state in H, and the intensity of the jump process at the state Φ is equal to $\varepsilon(\Phi, R^*R)$.

Now suppose that the system is controllable; specifically, let the energy E be a function of a parameter $u \in U$ whose value at each time can be chosen on the basis of the information τ available by this time. The opportunity to evaluate a posteriori states from Eq. (3.65) permits one to construct control strategies as functions of state, $u = u(t, \Phi(t))$. Suppose that we intend to maximize some operator-valued criterion of the form

$$\int_t^T (\Phi(s), A)\, ds + (\Phi(T), G), \tag{3.66}$$

where A and G are self-adjoint operators in H. Let $S(t, \Phi)$ denote the Bellman function, that is, the mathematical expectation (over all realizations of τ) of the maximum income of a process that starts at time t in a state Φ and terminates at time T. The Itô quantum stochastic formula in conjunction with standard heuristics of dynamic programming leads to the following Bellman equation for the optimal income function $S(t, \Phi)$ [**?**]:

$$\begin{aligned} &\frac{\partial S}{\partial t} + \varepsilon\left(\Phi, R^*R, \left(S\left(\frac{R\Phi R^*}{(\Phi, R^*R)}\right) - S(\Phi)\right) + (\Phi, A)\right. \\ &\quad \left. + \max_{u\in U}\left(\operatorname{grad}_\Phi S, i[E(u), \Phi] + \frac{\varepsilon}{2}\,(R^*R\Phi + \Phi R^*R) - \varepsilon(\Phi, R^*R)\Phi\right)\right. = 0. \end{aligned} \tag{3.67}$$

However, the solution of this equation with the terminal condition $S_T(\Phi) = (\Phi(T), G)$ is not uniquely determined in the class of functions smooth almost everywhere; hence, a well-grounded theory of generalized solutions should employ additional considerations so as to yield the Bellman function. For example, generalized solutions can be defined as limits of discrete approximations (cf. Theorem 3.7). Consider the special case of the unitary evolution (3.65) modeling the interaction of an atom with a Bose field by exchanging photons with simultaneous transitions of the atom from one level to another. In this model, R is the annihilation operator in the atom. In what follows we restrict ourselves to the case of two-level atoms, which is of physical interest. Thus, let $H = \mathbb{C}^2$, and suppose that the operators $E(u_1, u_2)$ and R are given by the matrices

$$\begin{pmatrix} \varepsilon_1 & u_1 + iu_2 \\ u_1 - iu_2 & \varepsilon_2 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

(This dependence of energy on the control parameters $u_1 \in [-A_1, A_1]$ and $u_2 \in [-A_2, A_2]$ implies that an external force (magnetic field) can be introduced in the system.) Let $\varepsilon_2 > \varepsilon_1$, and let $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $V = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ be, respectively, the lower (vacuum) and the excited eigenstates of the energy $E(0,0)$. Suppose that we intend to maintain the atom in a maximally excited state (that is, as close as possible to the state V) on a long observation interval. Then the operator A in Eq. (3.66) must be chosen as the operator R^*R of projection on V. The considered model can be regarded as the simplest model of laser pumping.

The density matrix of vector states of a two-dimensional atom is usually represented by the polarization vector $P = (p_1, p_2, p_3) \in S$ according to the formulas

$$\Phi = \frac{1}{2}(\mathrm{Id} + P_\sigma) = \frac{1}{2}(\mathrm{Id} + p_1\sigma_1 + p_2\sigma_2 + p_3\sigma_3),$$

where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli matrices. It is easy to verify that the filtering equation (3.65) in terms of P has the form

$$dP + (u_1 K_{u_1} + u_2 K_{u_2} + \Delta K_\Delta + \varepsilon K_\varepsilon) P\, dt = (v - P)\, dN,$$

where $v = \{p_1 = p_2 = 0,\ p_3 = 1\}$, dN is the differential of the Poisson process with density $\frac{\varepsilon}{2}(1 - p_3)$ at the state P, $\Delta = \varepsilon_2 - \varepsilon_1$ is the difference between the energy levels of $E(0,0)$, and the vector fields K on the sphere are determined by the formulas

$$K_{u_1}(P) = (0, -2p_3, 2p_2), \qquad K_{u_2}(P) = (2p_3, 0, -2p_1),$$
$$K_\Delta(P) = (p_2, -p_1, 0), \qquad K_\varepsilon(P) = (\tfrac{1}{2}p_1p_3, \tfrac{1}{2}p_2p_3, -\tfrac{1}{2}(1 - p_3^2)).$$

The Bellman equation (3.67) acquires the form

$$\frac{\partial S}{\partial t} + \frac{\varepsilon}{2}(1 - p_3)(S(v) - S(p)) + \frac{1}{2}(1 - p_3) + \left(\frac{\partial S}{\partial p}, \Delta K_\Delta + \varepsilon K_\varepsilon\right)$$
$$+ \max_{u_1, u_2} \left(\frac{\partial S}{\partial p}, u_1 K_{u_1} + u_2 K_{u_2}\right) = 0.$$

It is easy to see that we are just in the situation of Theorems 3.7 and 3.8. Hence, one can find the average income A_ε per unit time in this process under a permanent optimal control in the first approximation. It is equal to $1 + \varepsilon h^0(v)$, where $h^0(v)$ is the income provided by the optimal transition from v to V neglecting the interaction; this income can be evaluated according to Pontryagin's maximum principle.

3.4. The Bellman Differential Equation and the Pontryagin Maximum Principle for Multicriteria Optimization Problems

Idempotent structures adequate to multicriteria optimization are various semirings of sets or (in another representation) semirings of functions with an idempotent analog of convolution playing the role of multiplication. Here we use the methods of idempotent analysis to derive the differential equation describing the continuous dynamics of Pareto sets in multicriteria optimization problems and to define and construct generalized solutions of this equation.

Let $\le$ denote the Pareto partial order on $\mathbb{R}^k$. For any subset $M \subset \mathbb{R}^k$, by $\mathrm{Min}(M)$ we denote the set of minimal elements of the closure of M in $\mathbb{R}^k$. Following [**?**], we introduce the class $P(\mathbb{R}^k)$ of subsets $M \subset \mathbb{R}^k$ whose elements are pairwise incomparable,

$$P(\mathbb{R}^k) = \{M \subset \mathbb{R}^k : \mathrm{Min}(M) = M\}.$$

Obviously, $P(\mathbb{R}^k)$ is a semiring with respect to the operations $M_1 \oplus M_2 = \mathrm{Min}(M_1 \cup M_2)$ and $M_1 \odot M_2 = \mathrm{Min}(M_1 + M_2)$; the neutral element $\mathbb{0}$ with respect to addition in this semiring is the empty set, and the neutral element with respect to multiplication is the set whose sole element is the zero vector in $\mathbb{R}^k$. It is also clear that $P(\mathbb{R}^k)$ is isomorphic to the semiring of *normal sets*, that is, closed subsets $N \subset \mathbb{R}^k$ such that $b \in N$ implies $a \in N$ for any $a \ge b$; the sum and the product of normal sets are defined as their usual union and sum, respectively. Indeed, if N is normal, then $\mathrm{Min}(N) \in P(\mathbb{R}^k)$; conversely, with each $M \in P(\mathbb{R}^k)$ we can associate the normalization $\mathrm{Norm}(M) = \{a \in \mathbb{R}^k \mid \exists\, b \in M : a \ge b\}$.

For an arbitrary set X, by $B(X, \mathbb{R}^k)$ (respectively, by $B(X, P(\mathbb{R}^k))$) we denote the set of mappings $X \to \mathbb{R}^k$ (respectively, $X \to P(\mathbb{R}^k)$) bounded below with respect to the order. Then $B(X, P(\mathbb{R}^k))$ is a semimodule with respect to pointwise addition $\oplus$ and multiplication $\odot$ by elements of $P(\mathbb{R}^k)$.

These semirings and semimodules naturally arise in multicriteria dynamic programming problems. Let a mapping $f\colon X \times U \to X$ specify a controlled dynamical system on X, so that any choice of $x_0 \in X$ and of a sequence of controls $\{u_1, \dots, u_k\}$, $u_j \in U$, determines an admissible trajectory $\{x_j\}_{j=0}^k$ in X, where $x_j = f(x_{j-1}, u_j)$. Let $\varphi \in B(X \times U, \mathbb{R}^k)$, and let $\Phi(\{x_j\}_{j=0}^k) = \sum_{j=1}^k \varphi(x_{j-1}, u_j)$ be the corresponding vector criterion on the set of admissible trajectories. The element $\mathrm{Min}(\bigcup_{\{x_j\}} \Phi(\{x_j\}_{j=0}^k))$, where $\{x_j\}_{j=0}^k$ are all possible k-step trajectories issuing from x_0, is denoted by $\omega_k(x_0)$ and is called the *Pareto set* for the criterion Φ and the initial point x_0. Let us define the *Bellman operator* $\mathcal{B}$ on the semimodule $B(X, P(\mathbb{R}^k))$ by setting

$$(\mathcal{B}\omega)(x) = \mathrm{Min}\left(\bigcup_{u \in U} (\varphi(x, u) \odot \omega(f(x, u)))\right). \tag{3.68}$$

Obviously, $\mathcal{B}$ is linear in the semimodule $B(X, P(\mathbb{R}^k))$, and it follows from

Bellman's optimality principle that the Pareto sets in k-step optimization problems satisfy the recursion relation $\omega_k(x) = \mathcal{B}\omega_{k-1}(x)$.

Sometimes, it is convenient to use a different representation of the set $P(\mathbb{R}^k)$. Proposition 3.4 below, which describes this representation, is a specialization of a more general result stated in [**?**]. Let L denote the hyperplane in $\mathbb{R}^k$ determined by the equation

$$L = \Big\{(a^j) \in \mathbb{R}^k : \sum a^j = 0\Big\},$$

and let $CS(L)$ denote the semiring of functions $L \to \mathbb{R}\cup\{+\infty\}$ with pointwise minimum as addition and idempotent convolution

$$(g \circledast h)(a) = \inf_{b\in L}(g(a-b) + h(b))$$

as multiplication. Let us define a function $n \in CS(L)$ by setting $n(a) = \max_j(-a^j)$. Obviously, $n \circledast n = n$; that is, n is a multiplicatively idempotent element of $CS(L)$. Let $CS_n(L) \subset CS(L)$ be the subsemiring of functions h such that $n \circledast h = h \circledast n = h$. It is easy to see that $CS_n(L)$ contains the function identically equal to $\mathbb{0} = \infty$ and that the other elements of $CS_n(L)$ are just the functions that take the value $\mathbb{0}$ nowhere and satisfy the inequality $h(a) - h(b) \le n(a-b)$ for all $a, b \in L$. In particular, for each $h \in CS_n(L)$ we have

$$|h(a) - h(b)| \le \max_j |a^j - b^j| = \|a - b\|,$$

which implies that h is differentiable almost everywhere.

Proposition 3.4 *The semirings $CS_n(L)$ and $P(\mathbb{R}^k)$ are isomorphic.*

Proof. The main idea is that the boundary of each normal set in $\mathbb{R}^k$ is the graph of some real-valued function on L, and vice versa. More precisely, let us consider the vector $e = (1, \dots, 1) \in \mathbb{R}^k$ normal to L and assign a function $h_M\colon L \to \mathbb{R}$ to each set $M \in P(\mathbb{R}^k)$ as follows. For $a \in L$, let $h_M(a)$ be the greatest lower bound of the set of $\lambda \in \mathbb{R}$ for which $a + \lambda e \in \mathrm{Norm}(M)$. Then the functions corresponding to singletons $\{\varphi\} \subset \mathbb{R}^k$ have the form

$$h_\varphi(a) = \max_j(\varphi^j - a^j) = \overline{\varphi} + n(a - \varphi_L), \tag{3.69}$$

where $\overline{\varphi} = k^{-1}\sum_j \varphi^j$ and $\varphi_L = \varphi - \overline{\varphi}e$ is the projection of φ on L. Since idempotent sums $\oplus$ of singletons in $P(\mathbb{R}^k)$ and of functions (3.69) in $CS_n(L)$ generate $P(\mathbb{R}^k)$ and $CS_n(L)$, respectively, we can prove the proposition by verifying that the $\odot$-multiplication of vectors in $\mathbb{R}^k$ passes into the convolution of the corresponding functions (3.69). Namely, let us show that

$$h_\varphi \circledast h_\psi = h_{\varphi\oplus\psi}.$$

Indeed, by virtue of (3.69), it suffices to show that

$$n_\varphi \circledast n_\psi = n_{\varphi \oplus \psi},$$

where $n_\varphi(a) = n(a - \varphi_L)$, and the latter identity is valid since

$$n_\varphi \circledast n_\psi = n_0 \circledast n_{\varphi+\psi} = n \circledast n_{\varphi+\psi} = n_{\varphi+\psi}.$$

Finally, let us consider the controlled process in $\mathbb{R}^n$ specified by a controlled differential equation $\dot{x} = f(x,u)$ (where u belongs to a metric control space U) and by a continuous function $\varphi \in B(\mathbb{R}^n \times U, \mathbb{R}^k)$, which determines a vector-valued integral criterion

$$\Phi(x(\,\cdot\,)) = \int_0^t \varphi(x(\tau), u(\tau))\, d\tau$$

on the trajectories. Let us pose the problem of finding the Pareto set $\omega_t(x)$ for a process of duration t issuing from x with terminal set determined by some function $\omega_0 \in B(\mathbb{R}^n, \mathbb{R}^k)$, that is,

$$\omega_t(x) = \operatorname{Min} \bigcup_{x(\cdot)} (\Phi(x(\,\cdot\,)) \odot \omega_0(x(t))), \tag{3.70}$$

where $x(\,\cdot\,)$ ranges over all admissible trajectories issuing from x. By Proposition 3.4, we can encode the functions $\omega_t \in B(\mathbb{R}^n, P\mathbb{R}^k)$ by the corresponding functions

$$S(t,x,a)\colon \mathbb{R}_+ \times \mathbb{R}^n \times L \to \mathbb{R}.$$

The optimality principle permits us to write out the following equation, which is valid modulo $O(\tau^2)$ for small τ:

$$S(t,x,a) = \operatorname*{Min}_u (h_{\tau\varphi(x,u)} \circledast S(t-\tau, x + \Delta x(u)))(a). \tag{3.71}$$

It follows from the representation (3.69) of $h_{\tau\varphi(x,u)}$ and from the fact that n is, by definition, the multiplicative unit in $CS_n(L)$ that

$$S(t,x,a) = \min_u (\tau\overline{\varphi}(x,u) + S(t-\tau, x+\Delta x(u), a - \tau\varphi_L(x,u))).$$

Let us substitute $\Delta x = \tau f(x,u)$ into this equation, expand S in a series modulo $O(\tau^2)$, and collect similar terms. Then we obtain the equation

$$\frac{\partial S}{\partial t} + \max_u \left(\varphi_L(x,u)\,\frac{\partial S}{\partial a} - f(x,u)\,\frac{\partial S}{\partial x} - \overline{\varphi}(x,u) \right) = 0. \tag{3.72}$$

Although the presence of a vector criterion has resulted in a larger dimension, this equation coincides in form with the usual Bellman differential equation. Consequently, the generalized solutions can be defined on the basis of the idempotent superposition principle, as in §3.2. We thus obtain the main result of this section.

Theorem 3.9 *The Pareto set* $\omega_t(x)$ (3.70) *is determined by a generalized solution* $S_t \in B(\mathbb{R}^n, CS_n(L))$ *of Eq.* (3.72) *with the initial condition*

$$S_0(x) = h_{\omega_0(x)} \in B(\mathbb{R}^n, CS_n(L)).$$

The mapping $R_{CS}: S_0 \mapsto S_t$ *is a linear operator on* $B(\mathbb{R}^n, CS_n(L))$.

Note that $B(\mathbb{R}^n, S_n(L))$ is equipped with the $CS_n(L)$-valued bilinear inner product

$$\langle h, g\rangle = \inf_x (h \circledast g)(x).$$

By using the method of adjoint operator with respect to this product, one can prove (as was done in §3.2 for the usual Bellman equation) that the generalized solutions of Eq. (3.72) are the weak limits as $\tau \to 0$ of solutions of the time-discrete equation (3.71), the latter solutions being determined by iterations of the linear operator (3.69). This consideration can be used, in particular, for solving Eq. (3.72) numerically.

The application of Pontryagin's maximum principle to the problem in question is based on the following observation. Let R be the usual resolving operator for generalized solutions of the Cauchy problem for Eq. (3.72), so that R acts on the space $B(\mathbb{R}^n \times L, \mathbb{R} \cup \{+\infty\})$ of $\mathbb{R} \cup \{+\infty\}$-valued functions bounded below on $\mathbb{R}^n \times L$. Obviously, there is an embedding

$$\text{in}: B(\mathbb{R}^n, CS_n(L)) \to B(\mathbb{R}^n \times L, \mathbb{R} \cup \{+\infty\}),$$

which is an idempotent group homomorphism, that is, preserves the operation $\oplus = \min$. The diagram

$$\begin{array}{ccc} B(\mathbb{R}^n, CS_n(L)) & \xrightarrow{R_{CS}} & B(\mathbb{R}^n, CS_n(L)) \\ \downarrow \text{in} & & \downarrow \text{in} \\ B(\mathbb{R}^n \times L, \mathbb{R} \cup \{+\infty\}) & \xrightarrow{R} & B(\mathbb{R}^n \times L, \mathbb{R} \cup \{+\infty\}) \end{array}$$

commutes. Indeed, for smooth initial data this is obvious from the fact that a smooth solution of Eq. (3.72) always defines optimal synthesis. However, this implies commutativity for general initial conditions, since the operators R_{CS} and R are uniquely defined by their action on smooth functions and by the property that they are homomorphisms of the corresponding idempotent semigroups, that is, preserve the operation $\oplus = \min$. Thus, we have proved the following assertion.

Corollary 3.2 $S(t, x, a)$ *is the minimum of the functional*

$$\int_0^t \overline{\varphi}(x(\tau), u(\tau))\, d\tau + h_{\omega_0(x(t))}(a(t)) \tag{3.73}$$

defined on the trajectories of the system

$$\dot{x} = f(x, u), \qquad \dot{a} = -\varphi_L(x, u) \tag{3.74}$$

in $\mathbb{R}^n \times L$ *issuing from* (x, a) *with free right endpoint and fixed time* t.

Let us state a similar result for the case in which the time is not fixed. Namely, the problem is to find the Pareto set

$$\omega(x) = \operatorname{Min} \bigcup_{x(\cdot)} \Phi(x(\,\cdot\,)), \tag{3.75}$$

where Min is taken over the set of all trajectories of the equation $\dot{x} = f(x, u)$ joining a point $x \in \mathbb{R}^n$ with a given point ξ. For the corresponding function $S(x, a)$, we now obtain the stationary Bellman equation

$$\max_u \left(\varphi_L \frac{\partial S}{\partial a} - f(x, u) \frac{\partial S}{\partial x} - \overline{\varphi}(x, u) \right) = 0. \tag{3.76}$$

By analogy with the preceding case, we obtain the following assertion.

Corollary 3.3 *The Pareto set* (3.75) *is determined* (*by virtue of the isomorphism in Proposition* 3.4) *by the function* $S\colon \mathbb{R}^n \times L \to \mathbb{R}$ *such that* $S(x, a)$ *is the infimum of the functional*

$$\int_0^t \overline{\varphi}(x, u)\, d\tau + n(a(t))$$

defined on the trajectories of system (3.74) *issuing from* (x, a) *and satisfying the boundary condition* $x(t) = \xi$.

Example 3.1 Let us show how the method proposed works in the simplest variational problem with fixed endpoints and with two quadratic Lagrangians (quality criteria). Consider the following problem: Find the curves $x(\tau)$ in $\mathbb{R}$ joining the points $0 \in \mathbb{R}$ and $\xi \in \mathbb{R}$ and minimizing (in Pareto's sense) the integral functionals

$$\Phi_j = \int_0^t L_j(x(\tau), \dot{x}(\tau))\, d\tau, \qquad j = 1, 2,$$

with quadratic Lagrangians $L_j(x, v) = A_j x + xv + \frac{1}{2}v^2$, where A_j, $j = 1, 2$, are some constants (to be definite, we assume $A_1 > A_2$).

For the case of two criteria, the hyperplane L is a line parametrized by a single number, namely, by the first coordinate on $\mathbb{R}^2$; we have $n(a) = |a|$ for $a \in L$. The auxiliary system (3.74) in our case has the form

$$\dot{x} = u, \quad u \in \mathbb{R}, \qquad \dot{a} = -\frac{L_1 - L_2}{2} = \frac{A_2 - A_1}{2}\, x.$$

The corresponding Hamiltonian (Pontryagin) function has the form

$$H(x, a, p, \psi) = \max_u \left(pu + \psi\, \frac{L_2 - L_1}{2} - \frac{L_1 + L_2}{2} \right),$$

where ψ is the dual variable associated with a. The transversality condition on the right edge means that $\psi = 1$ if $a(t) < 0$, $\psi = -1$ if $a(t) > 0$, and $\psi \in [-1, 1]$ if $a(t) = 0$ (recall that ψ is constant along the trajectory, since a does not explicitly occur on the right-hand side in system (3.74)). Thus, we must solve a two-point variational problem with Lagrangian L_1 in the first case, L_2 in the second case, and $(L_1 + L_2)/2 + \psi(L_1 - L_2)/2$ in the third case, where ψ is determined from the condition $a(t) = 0$. These problems are uniquely solvable, and the solutions are given by quadratic functions of time. By explicitly finding these trajectories and by calculating

$$S(a,t) = \int_0^t \frac{L_1 + L_2}{2}\,(x(\tau), \dot{x}(\tau))\,d\tau + |a(t)|$$

along these trajectories, we obtain

$$1)\quad S(a,t) = \frac{1}{2}\,A_1\xi t - \frac{A_1^2 t^3}{24} + \frac{\xi^2(t+1)}{2t} - a \quad \text{if } \frac{24a}{(A_1 - A_2)t} \le 6\xi - A_1 t^2;$$

$$2)\quad S(a,t) = \frac{1}{2}\,A_2\xi t - \frac{A_2^2 t^3}{24} + \frac{\xi^2(t+1)}{2t} + a \quad \text{if } \frac{24a}{(A_1 - A_2)t} \ge 6\xi - A_2 t^2;$$

$$3)\quad S(a,t) = \frac{A_1 + A_2}{A_1 - A_2}\,a - \frac{12\xi a}{t^2(A_1 - A_2)} + \frac{24a^2}{t^3(A_1 - A_2)^2} + \frac{\xi^2(t+4)}{2t} \quad \text{if } 6\xi - A_1 t^2 \le \frac{24a}{(A_1 - A_2)t} \le 6\xi - A_2 t^2.$$

The Pareto set is given by 3) (the parts 1) and 2) give the boundary of its normalization). It is a segment of a parabola. For example, for $A_1 = 6$, $A_2 = 0$, and $\xi = t = 1$ we obtain $S(a,t) = \frac{2}{3}a^2 - a + \frac{5}{2}$, where $a \in [0, \frac{3}{2}]$. Carrying out the rotation according to Proposition 3.4, we see that the Pareto set is a segment of a parabola with endpoints $(2.5; 2.5)$ and $(4, 1)$.

3.5. A Boundary Value Problem for the Hamilton–Jacobi–Bellman Equation

In §3.2 we defined solutions of the stationary Bellman equation in the entire Euclidean space $\mathbb{R}^n$ as the eigenfunctions of the resolving operator for the corresponding Cauchy problem. In this section, we define generalized solutions of a boundary value problem for the equation

$$H(x, Dy) = 0 \tag{3.77}$$

in a bounded domain $X \subset \mathbb{R}^n$, where $Dy = \partial y/\partial x$ is the gradient of the unknown function. Our exposition follows the paper [**?**] (also see [**?**, **?**]. Other approaches to the solution of this problem can be found in [**?**, **?**, **?**, **?**]. The idea of our approach is to use algebraic structures of idempotent analysis so as to define semimodules of distributions and the differentiation operator in these semimodules in a way such that the solution of Eq. (3.77) could be defined as an element of this semimodule which turns Eq. (3.77) into an identity.

Let $X \subset \mathbb{R}^n$ be a closed domain, and let $B(X)$ be the semimodule of bounded mappings of X into $A = \mathbb{R} \cup \{+\infty\}$ (recall that the boundedness for A-valued functions is treated as the usual boundedness below). The inner product introduced in §1.2 is defined on $B(X)$. Let Φ be the set of continuously differentiable functions on X that tend to $\mathbb{0} = +\infty$ as the argument tends to the boundary Γ of the domain X.

Definition 3.2 [equivalence] Two functions $f, g \in B(X)$ are *equivalent* ($f \sim g$) if $(f, \varphi) = (g, \varphi)$ for any $\varphi \in \Phi$.

The inner product $(\,\cdot\,,\,\cdot\,)$ allows us to introduce the following weak convergence.

Definition 3.3 [weak convergence] A sequence of functions (equivalence classes) $\{f_i\}_{i=1}^{\infty}$ *weakly converges* to a function $f \in B(X)$ (we write $f = w_0\text{-}\lim f_i$) if for every $\varphi \in \Phi$ we have

$$\lim_{i\to\infty} (f_i, \varphi) = (f, \varphi).$$

Each equivalence class f contains a unique lower semicontinuous representative $\bar{f}$. If $\bar{f}$ is bounded above, then the equivalence class $-f$ is uniquely defined as the class containing the function $-\bar{f}$. Generally speaking, $-(-f) \neq f$.

Definition 3.4 [the space $P^0(X)$] The space $P^0(X)$ is the set of equivalence classes f (of functions from $B(X)$) such that $-f$ exists and $f = -(-f)$; we equip $P^0(X)$ with the weak convergence.

Since different continuous functions belong to different equivalence classes, we see that the following natural embedding is defined (and continuous)

$$i_0 \colon C(X) \to P^0(X).$$

Let ξ be a vector field on X, and let U be a differentiable coordinate neighborhood with coordinates $(x_1, \dots, x_n)$. Then ξ is given by a collection $(\xi_1, \dots, \xi_n)$ of scalar functions $\xi \colon U \to \mathcal{R}$.

Definition 3.5 [equivalent vector fields] Vector fields ξ and η on X are said to be *equivalent* if for any local coordinate neighborhood U the components ξ_i and η_i of ξ and η in U are equivalent for each $i = 1, \dots, n$.

Definition 3.6 [the space $P^0(TX)$ of vector fields on X] An element of $P^0(TX)$ is an equivalence class of vector fields for which in every local coordinate neighborhood U the components belong to $P^0(U)$. The convergence is defined as the componentwise convergence in $P^0(U)$ for every coordinate neighborhood U.

Example 3.2 Consider the sequence of functions $f_n = (1/n)\sin nx$ on some interval I in $\mathbb{R}$. This sequence converges in $P^0(I)$ to the function 0. The derivatives $f_n' = \cos nx$ converge in $P^0(I)$ to the function (equivalence class) identically equal to -1. But the sequence of vector fields on I given by the gradients Df_n is not convergent in $P^0(TI)$. Indeed, in the coordinate system on I induced by the standard coordinate in $\mathbb{R}$, the gradient Df_n coincides with $f_n' = \cos nx$ and converges to the function $x \to -1$, which defines a vector field ξ. In the coordinate system on I given by the mapping $x \to -x$ from I to $\mathbb{R}$ (change of direction in $\mathbb{R}$), Df_n are equal to $-\cos nx$ and converge to the function $x \to -1$, which defines the vector field $-\xi$. This example illustrates the natural requirement that the convergence of vector fields must be independent of the choice of a coordinate system. When the coordinate system changes, the limits of convergent sequences must change according to the transformation rules for vector fields, i.e., be multiplied by the Jacobi matrix of the coordinate transformation. It is useful to take into account this remark in the following definition.

Definition 3.7 [weak P'-convergence of scalar functions] A sequence of functions f_i (differentiable on X) P'-*converges* to a function (equivalence class) $f \in B(x)$ (we write $f = w_1\text{-lim}\, f_i$) if $f = w_0\text{-lim}_{i\to\infty} f_i$ and if the vector field $w_0\text{-lim}\, Df_i$ exists and belongs to $P^0(TX)$.

Proposition 3.5 *Suppose that* $w_1\text{-lim}\, f_i = w_1\text{-lim}\, g_i = f$; *then we have* $w_0\text{-lim}\, Df_i = w_0\text{-lim}\, Dg_i$, *i.e., the* w_0-lim *of the gradients depends only on* f *but not on the choice of a sequence that approximates* f.

Proof. First, let f be differentiable. If $F = w_1\text{-lim}\, Df_i$ and

$$F \neq (\partial f/\partial x_1, \dots, \partial f/\partial x_n),$$

then on some open set U and for some j we have $F_j(x) \neq (\partial f/\partial x_j)(x)$, $x \in U$, and the function $F_j - \partial f/\partial x_j$ does not change sign on this set; this contradicts the assumption that $w_0\text{-lim}\, f_i = f$.

Now let us consider the general case. If $F_j = w_0\text{-lim}\, \partial f_i/\partial x_j$, $G_j = w_0\text{-lim}\, \partial g_i/\partial x_j$, $F_j \neq G_j$, and $\partial f_i/\partial x_j < \partial g_i/\partial x_j$ for some j, then one can choose a differentiable function φ such that

$$\frac{\partial f}{\partial x_j} < \frac{\partial \varphi}{\partial x_j} < \frac{\partial g}{\partial x_j}$$

on some open set U. This immediately leads to a contradiction in just the same way as in the case of a differentiable function f.

By virtue of Proposition 3.5, for functions (equivalence classes) f that are P^1-limits of sequences of differentiable functions, the *gradient* Df is uniquely defined (as a class of equivalent vector fields).

Definition 3.8 [(the space $P^1(X)$ of scalar functions)]A function f (equivalence class) belongs to the space $P^1(X)$) if it is the w_1-limit of a sequence of differentiable functions. The convergence in the semiring $P^1(X)$ is determined by P^1-convergence in the sense of Definition 3.7, where the operation D can be applied to any elements from $P^1(X)$.

Proposition 3.6 *Each equivalence class* $\{f\} \in P^1(X)$ *contains a unique continuous function, which will be denoted by* f.

The proof is obvious.

In the sense of this assertion, we say that the elements of $P^1(X)$ are continuous functions. In particular, their restrictions to the boundary Γ of X are defined and continuous; this is important in what follows.

It is clear that any differentiable function $f: X \to \mathbb{R}$ determines an equivalence class in $P^1(X)$, so that there is a continuous embedding

$$i_1: C^1(X, \mathbb{R}) \to P^1(X).$$

Proposition 3.7 *The diagram*

$$\begin{array}{ccc} C^1(X,\mathbb{R}) & \stackrel{i_1}{\longrightarrow} & P^1(X) \\ \downarrow D & & \downarrow D \\ C(TX) & \stackrel{i_0}{\longrightarrow} & P^0(TX) \end{array}$$

commutes. Here $C(TX)$ *is the space of continuous vector fields equipped with uniform convergence, and the operators* D *denote the ordinary gradient* (*left vertical arrow*) *or the gradient* D *defined above* (*right vertical arrow*) *of a function from* $P^1(X)$.

The proof is obvious. (Note that the mapping $D: P^1 \to P^0$ is continuous.)

The following assertion is useful in the sequel.

Proposition 3.8 *Let* $\{f_i\}$ $(i \geq 1)$ *be a sequence of differentiable functions* w_0*-convergent to a function* $f \in P^0(X)$. *Furthermore, suppose that in some coordinate neighborhood* $(U, (x_1, \dots, x_n))$ *for every* $j = 1, \dots, n$ *the sequence* $\partial f_i/\partial x_j$ w_0*-converges to* $F_j \in P^0(X)$ *and the sequence* $-\partial f_i/\partial x_j$ w_0*-converges to* $G_j \in P^0(X)$. *Then for the sequence* $\{f_i\}$ *to* w_1*-converge to* f *in* U, *it is necessary and sufficient that* $F_j = -G_j$ *for every* $j = 1, \dots, n$ (*in the sense of elements of* P^0, *i.e., equivalence classes*).

Proof. Necessity is obvious. To prove the sufficiency, we note that if for each of the two w_0-convergent sequences a_i and b_i the upper and lower semicontinuous enveloping curves of the limits determine the same equivalence classes (i.e., elements from $P^0(X)$), then any linear combination $\{\lambda_1 a_i + \lambda_2 b_i\}$ $(\lambda_1, \lambda_2 \in \mathbb{R})$ also possesses the same property. Consequently, the application of the Jacobi matrix of the change of coordinates transforms the limits according to the transformation law for the principal part of vector fields, which proves the assertion.

The space $P^0(X)$ is a topological semimodule over the semiring $\mathcal{R}$, i.e., for any functions (equivalence classes) f_1 and f_2 from $P^0(X)$ and any $\lambda \in A$, the functions $f_1 \oplus f_2$ and $\lambda \odot f_1$ belong to P^0 (since these operations are defined pointwise on representatives of equivalence classes, it can be easily seen that these operations are well defined), and the operations of addition $\oplus$ and multiplication $\odot$ by λ determine continuous mappings $P^0 \times P^0 \to P^0$ and $P^0 \times \mathcal{R} \to P^0$, respectively. (Note, however, that the operation of ordinary addition $+$ is not continuous: $\sin^2 nx \to 0$ and $\cos^2 nx \to 0$ as $n \to \infty$, but $\sin^2 nx + \cos^2 nx = 1$.)

The space $P^1(X)$ is an (algebraic) semimodule over A, i.e., $f \oplus g \in P^1$, and $\lambda \odot f \in P^1$ for f, $g \in P^1$ and $\lambda \in A$, but the operation $\oplus$ is not continuous, as one can see from the following example. Let $f \equiv 0$, and let $g_\varepsilon(x) = |x| - \varepsilon$. We have $D(f \oplus g_0) = D(f) \equiv 0$, but

$$w_0\text{-}\lim_{\varepsilon\to 0}(D(f \oplus g_\varepsilon)) \not\equiv 0.$$

Proposition 3.9 *There exists a maximal* (*with respect to inclusion*) *subsemimodule of $P^1(X)$ containing all differentiable functions* (*i.e., elements from $i_1C^1(X,\mathbb{R})$*), *in which the operations of addition $\oplus$ and multiplication $\odot$ by elements of $\mathcal{R}$ are continuous in the topology induced from P^1. A function $f \in P^1(X)$ belongs to this subsemimodule if and only if the following property holds*: *suppose that for some $x_0 \in X$ and some differentiable nonnegative function $\varphi \in i_1C^1(X,\mathbb{R})$ we have $(f,\varphi) = f(x_0)$; then f is differentiable at the point x_0* (*here differentiability is understood in the classical sense*). *We denote this semimodule by $\mathcal{A}(X)$.*

Proof. The property mentioned in Proposition (3.7) can be restated as follows. If φ is a differentiable function such that $f - \varphi$ attains its minimum at the point x_0, then f is differentiable at x_0. Let $f \in \mathcal{A}$, and suppose that for some differentiable function φ the function $f - \varphi$ attains its minimum at a point x_0. Then f is differentiable at x_0, since otherwise the sequence $f \oplus \lambda_i \odot \varphi$ would not be w_1-convergent to $f \oplus \lambda \odot \varphi$, where λ satisfies $\lambda \odot \varphi(x_0) = f(x_0)$ and $\lambda_i \to \lambda$, $\lambda_i > \lambda$.

Conversely, suppose that the existence of a differentiable function φ such that $f - \varphi$ attains its minimum at x_0 implies the differentiability of f at x_0. Assume that $f \notin \mathcal{A}$; thus, there exists a sequence φ_i such that φ_i is w_1-convergent to φ but $f \oplus \varphi_i$ does not converge to $f \oplus \varphi$ in P^1. One can assume the functions φ_i and φ to be differentiable. This implies the existence of a point x_0 at which the function $f - \varphi$ attains its minimum; otherwise, small P^1-perturbations of φ do not remove jump changes of the derivatives of $f \oplus \varphi$. But then f is differentiable at the point x_0, and so it is impossible for $f \oplus \varphi_i$ not to converge in P^1 to $f \oplus \varphi$ for smooth φ and φ_i.

Example 3.3 The function $x \to |x|$ does not belong to $\mathcal{A}(I)$, where I is a an interval in $\mathbb{R}$ containing 0, but the function $x \to -|x|$ belongs to $\mathcal{A}(I)$.

Let H be a mapping of $X \otimes \mathbb{R}^n$ into $\mathbb{R}$ (the Hamiltonian), which will be assumed continuous. We consider Eq. (3.77), where $y \in P^1(X)$, Dy is the gradient of the function y in the sense described above, and the equality is understood in the sense of an equality in $P^0(X)$, i.e., as the coincidence of equivalence classes. Thus, we consider the mapping of $P^1(X)$ into $P^0(X)$ determined by the function H and denoted by $\mathcal{H}$ in what follows:

$$\mathcal{H}(y)(x) = H(x, Dy(x)).$$

Our basic assumption is that $\mathcal{H}$ is continuous.

Example 3.4 Let X be an interval in $\mathbb{R}$, and let $H(x,p) = -p^2 + a(x)$, where $a \in C(X, \mathbb{R})$. The mapping $\mathcal{H}\colon y \to -(y_x')^2 + a(x)$ of P^1 into P^0 is continuous. On the other hand, the mapping $\mathcal{H}_1$ determined by the function $H(x,p) = p^2$ is not continuous. Indeed, let $g(x) = -|x|$, and let $g_n(\,\cdot\,)$ be any sequence of differentiable functions convergent to g in the topology of P^1, for example,

$$g_n(x) = \begin{cases} -|x|, & |x| \geq n^{-1}, \\ -nx^2/2 - n^{-1}/2, & |x| \leq n^{-1}. \end{cases}$$

Then

$$\mathcal{H}(g) = w_0\text{-}\lim \mathcal{H}(g_n) = -1, \qquad \mathcal{H}_1(g) = 1,$$

but

$$\mathcal{H}_1(g_n) = \begin{cases} 1, & \text{for } |x| \geq n^{-1}, \\ n^2x^2, & \text{for } |x| < n^{-1}, \end{cases}$$

so that $w_0\text{-}\lim \mathcal{H}_1(g_n) \neq 1$.

Proposition 3.10 *Let $X \subset \mathbb{R}^n$. Suppose that the mapping $y \to \mathcal{H}(y)$ of $P^1(X)$ into $P^0(X)$, where $\mathcal{H}(y)(x) = H(x, Dy(x))$, is continuous. Then for any $x \in X$ and $\lambda \in \operatorname{Im} H(x, \cdot\,)$, the set*

$$\Lambda_{x,\lambda} = \{p \in \mathbb{R}^n \mid H(x,p) \geq \lambda\}$$

is a convex subset of $\mathbb{R}^n$.

Proof. Let a, $b \in \mathbb{R}^n$ and $x_0 \in X$. Let $\langle\,\cdot\,,\cdot\,\rangle$ be the Euclidean inner product on $\mathbb{R}^n$. Consider the function

$$\varphi_{a,b}(x) = \min\{\langle a, x\rangle + m, \langle b, x\rangle\},$$

where the number $m \in \mathbb{R}$ is chosen so that

$$\langle a, x_0\rangle + m = \langle b, x_0\rangle.$$

The function $\varphi_{a,b}$ belongs to the space $P^1(X)$, and a sufficiently small neighborhood U of the point x_0 is split by the hyperplane

$$\mathfrak{M}^0 = \{x \mid \langle a-b, x\rangle + m = 0\}$$

into the subset U^+ where $D\varphi = a$ and the subset U^- where $D\varphi = b$. Let φ_n be a sequence of differentiable functions on X w_1-convergent to $\varphi_{a,b}$. For any interior point z of the segment I in $\mathbb{R}^n$ that connects the points a and b, we have $z \in \operatorname{Im} D\varphi_n$ for n sufficiently large.

Let x_0 and λ be such that the set $\Lambda_{x_0,\lambda}$ contains at least two points. Assume, on the contrary, that there exist two different points a and b from $\Lambda_{x,\lambda}$ and a point z from the segment I joining a and b in $\mathbb{R}^n$ such that $z \notin \Lambda_{x_0,\lambda}$. Let $l = H(x_0, z)$. Then $l < \lambda$. We have $\mathcal{H}(\varphi_{a,b})(x_0) \geq \lambda$, whereas for the members φ_n of the sequence w_1-convergent to $\varphi_{a,b}$ there always exists a point x_n such that $\mathcal{H}(\varphi_n)(x_n) = l$ and $x_n \to x$ as $n \to \infty$. Thus, the sequence $\mathcal{H}(\varphi_n)$ does not converge in P^0 to $\mathcal{H}(\varphi_{a,b})$, which contradicts the continuity of $\mathcal{H}\colon P^1 \to P^0$. This contradiction proves that the set $\Lambda_{x_0,\lambda}$ is convex.

Remark 3.3 Usually, the convexity of the Hamiltonian $H(x, \cdot\,)$ is not necessary for the mapping $\mathcal{H}$ to be continuous. For example, if $X \subset \mathbb{R}$ and $H\colon \mathbb{R} \to \mathbb{R}$ does not depend on x, then the mapping $y \mapsto \mathcal{H}y = H(Dy)$ of P^1 into P^0 is continuous whenever H is monotone or has a unique local extremum that is a maximum.

We shall say that a subset $\Lambda \subset \mathbb{R}^n$ is *strictly convex* if all interior points of any segment that connects two different points of Λ are interior points of Λ.

Theorem 3.10 *Suppose that X is a bounded closed domain in $\mathbb{R}^n$ with boundary Γ and $H\colon X \times \mathbb{R}^n \to \mathbb{R}$ is a continuous mapping. Let the following conditions be satisfied*:

(1) *the mapping $\mathcal{H}\colon P^1(X) \to P^0(X)$, $(\mathcal{H}y)(x) = H(x, Dy)$, is continuous*;

(2) *the set $\Lambda_x = \{p \in \mathbb{R}^n \mid H(x,p) \geq 0\}$ is nonempty, compact, and strictly convex for all $x \in X$*;

(3) *there exists a function $\bar{y} \in P^1(X)$ such that*

$$H(x, D\bar{y}(x)) \geq 0. \qquad (3.78)$$

Then Eq. (3.77) *has a solution $y \in \mathcal{A}(X)$ such that the restrictions of y and $\bar{y}$ to the boundary Γ coincide.*

Proof. The restriction of $\bar{y}$ to Γ will be denoted by g. Let F be the pointwise supremum of all P^1-solutions f of the inequality $\mathcal{H}(f) \geq 0$ whose restrictions to Γ coincide with the function g, i.e.,

$$F(x) = \sup f(x).$$

The set of such functions f is nonempty by condition (3) of the theorem. One can show that the function F is a solution of the equation $\mathcal{H}(y) = 0$ (which may have more than one solution); see details in [**?**].

Definition 3.9 [stable solutions] A solution $y_0 \in \mathcal{A}(X)$ of Eq. (3.77) with the boundary condition $y_0|_\Gamma = g$ is said to be *stable* if there exists a $\lambda_0 < 0$ such that for every $\lambda \in [\lambda_0, 0]$ there exists a solution $y_\lambda \in \mathcal{A}(X)$ of the equation

$$H(x, Dy(x)) = \lambda \tag{3.79}$$

with boundary condition $y_\lambda|_\Gamma = g$ such that in the metric of $C(X, \mathbb{R})$ we have

$$y_0 = \lim_{\lambda \to 0, \lambda < 0} y_\lambda.$$

The construction of the solution described in Theorem 3.10 (as the supremum of P^1-solutions of inequality (3.78) satisfying the boundary condition) readily implies the following proposition.

Proposition 3.11 *Let condition* (2) *of Theorem* 3.10 *be replaced by the following condition*:

(2′) *the set* $\Lambda_{x,\lambda} = \{p \in \mathbb{R}^n \mid H(x,p) \geq \lambda\}$ *is nonempty, compact, and strictly convex for all* λ *from some interval* $(\lambda_0, 0]$, *where* $\lambda_0 < 0$.

Then there exists a stable solution $y_\lambda \in \mathcal{A}(X)$ *of Eq.* (3.79) *with the boundary condition* $y_\lambda|_\Gamma = \bar{y}|_\Gamma$ *for* $\lambda \in (\lambda_0, 0]$.

Remark 3.4 It follows from condition (2) of Theorem 3.10 that $\Lambda_{x,\lambda}$ is nonempty and compact for λ from some interval $(\lambda_0, 0]$, where $\lambda_0 < 0$.

Theorem 3.11 *Let the mapping*

$$\mathcal{H}: P^1(X) \to P^0(X), \qquad (\mathcal{H}y)(x) = H(x, Dy(x))$$

be continuous, and let the equation $\mathcal{H}y = 0$ *with boundary value* g *on* Γ *have a stable solution from* $\mathcal{A}(X)$. *Then this solution is the unique stable solution of the equation* $\mathcal{H}y = 0$ *with the boundary value* g *in the semimodule* $\mathcal{A}(\mathcal{X})$.

Proof. Assume that f_1 is stable and f_2 is any other (not necessarily stable) solution of Eq. (3.77) with boundary value g on Γ. Let h_λ be a solution of Eq. (3.79) with boundary value g on Γ that is close (in the metric of C) to the solution f_1 for λ close to 0, $\lambda_0 < \lambda < 0$.

Proposition 3.12 *Suppose that the mapping* $\mathcal{H}: P^1(X) \to P^0(X)$ *is continuous, a function* $f \in \mathcal{A}(X)$ *satisfies Eq.* (3.77), *a function* $h \in \mathcal{A}(X)$ *satisfies Eq.* (3.79) (*for* $\lambda < 0$), *and* $f|_\Gamma = h|_\Gamma$. *Then for all* $x \in X$ *we have* $h(x) \geq f(x)$.

Proof. Assume, on the contrary, that there exists a point $x \in X$ such that $f(x_0) > h(x_0)$. Since both functions belong to $P^1(X)$, we obtain

$$f(x) > h(x), \qquad x \in U, \tag{3.80}$$

in some neighborhood U of the point x_0 ($U \subset X$). Consider the function $S_\mu = f \oplus \mu \odot h = \min\{f(\cdot), \mu + h(\cdot)\}$, which belongs to the semimodule $\mathcal{A}(X)$ for all $\mu \in \mathbb{R}$. Next, consider the set

$$\mathcal{O}_\mu = \{x \in X \mid \mu + h(x) < f(x)\}.$$

For $\mu > 0$ this set has an empty intersection with Γ, since S_μ coincides with f in some neighborhood of Γ. For $\mu > 0$ sufficiently small, the set $\mathcal{O}_\mu$ is nonempty in view of inequality (3.80), whereas for μ sufficiently large, $\mathcal{O}_\mu$ is empty. Furthermore, $\mathcal{O}_{\mu_1} \subset \mathcal{O}_{\mu_2}$ for $\mu_1 > \mu_2$. Therefore, there exists a $\bar{\mu}$ such that for $\mu \geq \bar{\mu}$ the function S_μ coincides with f and for $\mu < \bar{\mu}$ the function S_μ coincides with $\mu \odot h$ on an open set. The sequence $S_i = f \oplus \mu_i \odot h$ of elements of $\mathcal{A}$ converges in P^1 to the function $S = f \oplus \bar{\mu} \odot h = f$ as $\mu_i \to \bar{\mu}$ ($\mu_i < \bar{\mu}$) by the definition of the semimodule $\mathcal{A}$. The continuity of the mapping $\mathcal{H}: P^1 \to P^0$ implies the convergence of the sequence $\mathcal{H}(S_i)$ to $\mathcal{H}(S) = \mathcal{H}(f)$. But this is not true, since $\mathcal{H}(S) = 0$, and for every i the function $\mathcal{H}(S_i)$ is equal to $\lambda < 0$ on the open set $\mathcal{O}_{\mu_i}$; hence, $w_0\text{-}\lim \mathcal{H}(S_i) \neq 0$. The contradiction obtained proves that $f(x) \leq h(x)$ for all $x \in X$.

To prove Theorem 3.11, it suffices to note that the stable solution, according to Proposition 3.12, always coincides with the supremum of all possible $\mathcal{A}$-solutions with the boundary value g and hence is unique.

Example 3.5 Let $X = [-1, 1]$, $g(1) = g(-1) = 0$, and $H(x, p) = -p^2 + x^4$. It is obvious that the equation

$$-(Dy)^2 + x^4 = 0 \tag{3.81}$$

has the differentiable (and hence belonging to $\mathcal{A}(X)$) solutions

$$\varphi_1: x \mapsto |x|^3/3 - 1/3 \quad \text{and} \quad \varphi_2: x \mapsto -|x|^3/3 + 1/3$$

with zero boundary values. Hence, for $\nu_1, \nu_2 \in \mathcal{R}$ such that $v_1 \oplus v_2 = \mathbb{1}$, the function $\nu_1 \odot \varphi_1 \oplus \nu_2 \odot \varphi_2$ is also a solution of Eq. (3.81) belonging to $\mathcal{A}$ and satisfying the boundary condition. Thus, there are infinitely many solutions in $\mathcal{A}(X)$ of the Dirichlet problem for this equation, but only the solution φ_1 is stable. Note that in the family of equations

$$-(Dy)^2 + x^4 = \lambda$$

for $\lambda < 0$ there exists a unique solution from $\mathcal{A}$ with zero boundary value, and for $\lambda > 0$ there exists no solution at all. This is a general law: nonstable solutions of the Dirichlet problem for Eq. (3.77) arise only for critical values of the parameters, whereas uniqueness and stability of the solution of the Dirichlet problem is a generic property.

Remark 3.5 If one regards Eq. (3.77) from the point of view of the papers [**?**, **?**, **?**], then the viscosity solutions [**?**, **?**] or the min-max solutions [**?**] coincide with those given by Theorems 3.10 and 3.11.

3.6. The Stochastic Hamilton–Jacobi–Bellman Equation

In this section we explain some results obtained in [**?**], where also the application of these results to the construction of WKB-type asymptotics of stochastic pseudodifferential equations is presented. Here we study the equation

$$dS + H\left(t, x, \frac{\partial S}{\partial x}\right) dt + \left(c(t,x) + g(t,x)\frac{\partial S}{\partial x}\right) \circ dW = 0, \tag{3.82}$$

where $x \in \mathcal{R}^n$, $t \geq 0$, $W = (W^1, \dots, W^n)$ is the standard n-dimensional Brownian motion ($\circ$, as usual, denotes the Stratonovich stochastic differential), $S(t, x, [W])$ is an unknown function, and the Hamiltonian $H(t, x, p)$ is convex with respect to p. This equation can naturally be called the stochastic Hamilton–Jacobi–Bellman equation. First, we explain how this equation appears in the theory of stochastic optimization. Then we develop the stochastic version of the method of characteristics to construct classical solutions of this equation, and finally, on the basis of the methods of idempotent analysis (and on analogy with the deterministic case), we construct a theory of generalized solutions of the Cauchy problem for this equation. Let the controlled stochastic dynamics be defined by the equation

$$dx = f(t, x, u)\, dt + g(t, x) \circ dW, \tag{3.83}$$

where the control parameter u belongs to some metric space U and the functions f and g are continuous in t and u and Lipschitz continuous in x. Let the income along the trajectory $x(\tau)$, $\tau \in [t, T]$, defined by the starting point $x = x(0)$ and the control $[u] = u(\tau)$, $\tau \in [t, T]$, be given by the integral

$$I_t^T(x, [u], [W]) = \int_t^T b(\tau, x(\tau), u(\tau))\, d\tau + \int_t^T c(\tau, x(\tau)) \circ dW. \tag{3.84}$$

We are looking for an equation for the cost (or Bellman) function

$$S(t, T, x, [W]) = \sup_{[u]} (I_t^T(x, [u], [W]) + S_0(x(T))), \tag{3.85}$$

where the supremum is taken over all piecewise smooth (or equivalently, piecewise constant) controls $[u]$ and S_0 is some given function (terminal income). Our argument is based on the following well-known fact: if we approximate the noise W in some stochastic Stratonovich equation by smooth functions of the form

$$W = \int q(\tau)\, d\tau, \tag{3.86}$$

with some continuous $q(\tau)$, then the solutions of the corresponding classical (deterministic) equations will tend to the solution of the given stochastic equation. Thus, for W of the form (3.86), we have the dynamics

$$\dot{x} = f(\tau, x, u) + g(\tau, x) q(\tau)$$

and the integral income

$$\int_t^T [b(\tau, x(\tau), u(\tau)) + c(\tau, x(\tau))q(\tau)]\, d\tau.$$

On writing out the Bellman equation for the corresponding deterministic (non-homogeneous) optimization problem, we obtain

$$\frac{\partial S}{\partial t} + \sup_u \left(b(t,x,u) + f(t,x,u)\frac{\partial S}{\partial x} \right) + \left(c(t,x) + g(t,x)\frac{\partial S}{\partial x} \right) q(t) = 0.$$

By rewriting this equation in the stochastic Stratonovich form, we obtain (3.82) with

$$H(t,x,p) = \sup_u (b(t,x,u) + pf(t,x,u)).$$

Let us consider the following two special cases.

(i) $c = 0$ and $g = g(t)$ is independent of x. Then, by differentiating (3.82), we obtain

$$d\,\frac{\partial S}{\partial x} + \left(\frac{\partial H}{\partial x} + \frac{\partial H}{\partial p}\,\frac{\partial^2 S}{\partial x^2} \right) dt + g\,\frac{\partial^2 S}{\partial x^2} \circ dW = 0,$$

and using the relationship

$$v \circ dW = v\, dW + \frac{1}{2}\, dv\, dW \tag{3.87}$$

between the Itô and the Stratonovich differentials, we obtain the equation for S in the Itô form

$$dS + H\left(t, x, \frac{\partial S}{\partial x} \right) dx + \frac{1}{2}\, g^2 \frac{\partial^2 S}{\partial x^2}\, dt + g\,\frac{\partial S}{\partial x}\, dW = 0. \tag{3.88}$$

For the mean optimal cost function $\widetilde{S}$, this implies the standard second-order Bellman equation of the stochastic control theory:

$$\frac{\partial \widetilde{S}}{\partial t} + H\left(t, x, \frac{\partial \widetilde{S}}{\partial x} \right) + \frac{1}{2}\, g^2 \frac{\partial^2 \widetilde{S}}{\partial x^2} = 0.$$

(ii) $g = 0$. Then Eq. (3.82) acquires the form

$$dS + H\left(t, x, \frac{\partial S}{\partial x} \right) dt + c(t,x)\, dW = 0, \tag{3.89}$$

since in this case the Itô and the Stratonovich differential forms coincide.

For simplicity, we shall study the case most important for applications in detail. Namely, we consider Eq. (3.89) with H and c that do not explicitly depend on t (see remarks on the general equation (3.82) in the end of the present section). Our main tool is the stochastic Hamiltonian system

$$\begin{cases} dx = \dfrac{\partial H}{\partial p}\, dt, \\ dp = -\dfrac{\partial H}{\partial x}\, dt - c'(x)\, dW. \end{cases} \tag{3.90}$$

The general theory of such systems (with Hamiltonians $H(x,p)$ quadratic in p), in particular, existence and uniqueness theorems for their solutions, can be found in [**?**].

Theorem 3.12 *For fixed $x_0 \in \mathcal{R}^n$ and $t > 0$, let us consider the map $P\colon p_0 \mapsto x(t,p_0)$, where $(x(\tau,p_0),\ p(\tau,p_0))$ is the solution of* (3.90) *with initial values (x_0,p_0). Suppose that all second derivatives of the functions H and c are uniformly bounded, the matrix $\mathrm{Hess}_p\, H$ of the second derivatives of H with respect to p is uniformly positive (i.e., $\mathrm{Hess}_p\, H \geq \lambda E$ for some constant λ), and for any fixed x_0 all matrices $\mathrm{Hess}_p\, H(x_0,p)$ commute. Then the map P is a diffeomorphism for small $t \leq t_0$ and all x_0.*

We postpone the rather technical proof of this theorem until the end of the section and first explain the main construction of classical and generalized solutions of Eq. (3.89). Let us define the two-point stochastic action

$$S_W(t,x,\xi) = \inf \int_0^t (L(q,\dot q)\, d\tau - c(q)\, dW), \tag{3.91}$$

where the infimum is taken over all piecewise smooth curves $q(\tau)$ such that $q(0) = \xi$ and $q(t) = x$, and the Lagrangian L is, as usual, the Legendre transform of the Hamiltonian H with respect to the last argument.

Theorem 3.13 *Suppose that the assumptions of Theorem* 3.12 *are satisfied and $t \leq t_0$. Then*

$$\text{(i)}\quad S_W(t,x,\xi) = \int_0^t (p\, dq - H(x,p)\, dt - c(x)\, dW), \tag{3.92}$$

where the integral is taken along the trajectory $(x(\tau),\ p(\tau))$ that joins the points ξ and x (and which exists by Theorem 3.12*);*

(ii) $p(t) = \dfrac{\partial S}{\partial x}, \quad p_0 = -\dfrac{\partial S}{\partial \xi};$

(iii) *S satisfies Eq.* (3.89) *as a function of x;*

(iv) *$S(t,x,\xi)$ is convex in x and ξ.*

Proof. As above, we use the approximation of the Wiener trajectories W by a sequence of smooth functions W_n of the form (3.86). For these functions, Eq. (3.89), as well as system (3.90), becomes classical and the results of the theorem are well known (e.g., see [**?**, **?**]). By the cited approximation theorem, the corresponding sequence of diffeomorphisms P_n in Theorem 3.12 converges to the diffeomorphism P, and moreover, by virtue of the uniform estimates of their derivatives (see formulas (3.99) and (3.100) below), the convergence of $P_n(t, p_0)$ to $P(t, p_0)$ is locally uniform, as well as the convergence of the inverse diffeomorphisms $P_n^{-1}(t, x) \to P^{-1}(t, x)$. This implies the convergence of the corresponding solutions S_n to the function (3.92) together with the x-derivatives. Again by the approximation argument, we conclude that the limit function satisfies Eq. (3.89). Let us also note that the convexity of S follows from the equations

$$\frac{\partial^2 S}{\partial x^2} = \frac{\partial p}{\partial x} = \frac{\partial p}{\partial x_0}\left(\frac{\partial x}{\partial x_0}\right)^{-1} = \frac{1}{t}\,(1 + O(t^{1/2})),$$
$$\frac{\partial^2 S}{\partial \xi^2} = -\frac{\partial p_0}{\partial \xi} = \left(\frac{\partial x}{\partial p_0}\right)^{-1}\frac{\partial x}{\partial \xi} = \frac{1}{t}\,(1 + O(t^{1/2})),$$

which can be obtained from (ii), formulas (3.99), (3.100) (see below), and again by the approximation argument.

By a similar argument, one proves the following theorem.

Theorem 3.14 *Let $S_0(x)$ be a smooth function, and suppose that for all $t \le t_0$ and $x \in \mathcal{R}^n$ there exists a unique $\xi = \xi(t, x)$ such that $x(t, \xi) = x$ for the solution $(x(\tau, \xi),\ p(\tau, \xi))$ of system* (3.90) *with the initial data $x_0 = \xi$, $p_0 = (\partial S_0/\partial x)(\xi)$. Then*

$$S(t, x) = S_0(\xi) + \int_0^t (p\,dq - H(x, p)\,dt - c(x)\,dW) \tag{3.93}$$

(*where the integral is taken along the trajectory $(x(\tau, \xi),\ p(\tau, \xi))$*) *is the unique classical solution of the Cauchy problem for Eq.* (3.89) *with the initial function $S_0(x)$.*

Remark 3.6 For Hamiltonians of the form $H = p^2 + V(x)$, this theorem was also obtained by A. Truman and H. Z. Zhao [**?**]. Their proof is different and is based on direct calculation using Itô's differential and Itô's formula.

Theorems 3.12 and 3.13, in fact, give simple sufficient conditions under which the assumptions of Theorem 3.14 are satisfied. The following result is a direct corollary of Theorem 3.13.

Theorem 3.15 *Let the assumptions of Theorem* 3.12 *hold, and let the function $S_0(x)$ be smooth and convex. Then for $t \le t_0$ there exists a unique classical* (*i.e., almost everywhere smooth*) *solution of the Cauchy problem for Eq.*

(3.89) *with the initial function* $S_0(x)$*. This solution is given by Eq.* (3.93)*, or, equivalently, by the formula*

$$R_t S_0(x) = S(t,x) = \min_{\xi}(S_0(\xi) + S_W(t,x,\xi)). \tag{3.94}$$

Now one can directly apply the method for constructing the generalized solution of the deterministic Bellman equation to the stochastic case, thus obtaining the following theorem.

Theorem 3.16 *For any initial function* $S_0(x)$ *bounded below, there exists a unique generalized solution of the Cauchy problem for Eq.* (3.89)*, which is given by* (3.93) *for all* $t \geq 0$*.*

Remark 3.7 Approximating nonsmooth Hamiltonians by smooth functions and defining the generalized solutions as the limits of the solutions corresponding to the smooth Hamiltonians, we find (on analogy with the deterministic case) that formula (3.94) for generalized solutions remains valid for nonsmooth Hamiltonians.

Proof of Theorem 3.12. Obviously, the solution of the linear matrix equation

$$dG = B_1 G\,dt + B_2(t)\,dW, \qquad G|_{t=0} = G_0, \tag{3.95}$$

where $B_j = B_j(t,[W])$ are given uniformly bounded and nonanticipating functionals on the Wiener space, can be represented by the convergent series

$$G = G_0 + G_1 + G_2 + \dots, \tag{3.96}$$

where

$$G_k = \int_0^t B_1(\tau) G_{k-1}(\tau)\,d\tau + \int_0^t B_2(\tau) G_{k-1}(\tau)\,dW(\tau). \tag{3.97}$$

By differentiating (3.90) with respect to the initial data (x_0, p_0), one finds that the matrix

$$G = \frac{\partial(x,p)}{\partial(x_0,p_0)} = \begin{pmatrix} \frac{\partial x}{\partial x_0} & \frac{\partial x}{\partial p_0} \\ \frac{\partial p}{\partial x_0} & \frac{\partial p}{\partial p_0} \end{pmatrix} (x(\tau,[W]), p(\tau,[W]))$$

satisfies the following special case of (3.95):

$$\begin{aligned} dG &= \begin{pmatrix} \frac{\partial^2 H}{\partial p \partial x} & \frac{\partial^2 H}{\partial p^2} \\ -\frac{\partial^2 H}{\partial x^2} & -\frac{\partial^2 H}{\partial x \partial p} \end{pmatrix} (x,p) G\,dt - \begin{pmatrix} 0 & 0 \\ c''(x) & 0 \end{pmatrix} G\,dW, \\ G_0 &= \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}. \end{aligned} \tag{3.98}$$

Let us write $f = \widetilde{O}(t^\alpha)$ if $f = O(t^{\alpha-\varepsilon})$ for any $\varepsilon > 0$ as $t \to 0$. Applying the log log law for stochastic integrals to the solutions of system (3.90) and then calculating G_1 by (3.97), we obtain

$$G_1 = \left(t \begin{pmatrix} \frac{\partial^2 H}{\partial p \partial x} & \frac{\partial^2 H}{\partial p^2} \\ -\frac{\partial^2 H}{\partial x^2} & -\frac{\partial^2 H}{\partial x \partial p} \end{pmatrix} (x_0, p_0) + \widetilde{O}(t^{3/2}) + \begin{pmatrix} 0 & 0 \\ c''(x_0)\widetilde{O}(t^{1/2}) & 0 \end{pmatrix} \right) \times \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}.$$

Again applying the log log law to the subsequent terms of the series (3.96), we readily find that the remainder $G - G_0 - G_1$ is $\widetilde{O}(t^{3/2})$. Thus, the series (3.96) for system (3.98) converges, and we obtain the following approximate formulas for the solutions:

$$\begin{aligned} \frac{\partial x}{\partial x_0} &= E + t\,\frac{\partial^2 H}{\partial p \partial x}(x_0, p_0) + \widetilde{O}(t^{3/2}), \\ \frac{\partial x}{\partial p_0} &= t\,\frac{\partial^2 H}{\partial p^2}(x_0, p_0) + \widetilde{O}(t^{3/2}), \end{aligned} \tag{3.99}$$

$$\begin{aligned} \frac{\partial p}{\partial x_0} &= \widetilde{O}(t^{1/2}), \\ \frac{\partial p}{\partial p_0} &= E + t\,\frac{\partial^2 H}{\partial x \partial p}(x_0, p_0) + \widetilde{O}(t^{3/2}). \end{aligned} \tag{3.100}$$

From these formulas, one concludes that the map $P\colon p_0 \mapsto x(t, p_0)$ is a local diffeomorphism. Let us prove that it is injective. Indeed, since

$$x(t, p_1) - x(t, p_2) = \int_0^t \frac{\partial x}{\partial p_0}\,(p_1 + \tau(p_2 - p_1))(p_2 - p_1)\,d\tau,$$

we have

$$\begin{aligned} (x(t, p_1) - x(t, p_2))^2 &= \int_0^1\int_0^1 \left(\left(\frac{\partial x}{\partial p_0}\right)^t (p_1 + s(p_2 - p_1)) \right. \\ &\quad \times \left. \frac{\partial x}{\partial p_0}\,(p_1 + \tau(p_2 - p_1))(p_2 - p_1), (p_2 - p_1) \right) d\tau\, ds \\ &\geq C\|p_2 - p_1\|^2. \end{aligned} \tag{3.101}$$

The last inequality is due to the second formula in (3.99) and to the properties of $\mathrm{Hess}_p\, H$ mentioned in the conditions of Theorem 3.12. It follows from the estimate (3.101) that the considered map is injective and, moreover, that $x(t, p_0) \to \infty$ as $p_0 \to \infty$. From this, one concludes that the image of this map is simultaneously closed and open and therefore coincides with the entire space. The proof of Theorem 3.12 is complete.

To conclude this section, let us note that the general equation (3.82) can be treated in a similar way. The difference is in that the corresponding Hamiltonian system will acquire the form

$$\begin{cases} dx = \dfrac{\partial H}{\partial p}\,dt + g(t,x)\circ dW, \\ dp = -\dfrac{\partial H}{\partial x}\,dt - c'(t,x)\circ dW \end{cases}$$

(thus, the Stratonovich form no longer coincides with the Itô form, as was the case for system (3.82)), and therefore, the estimates (3.99) and (3.100) must be replaced by less sharp estimates.

3.7. The Turnpike Theorem in General Topological Spaces and a Limit Theorem for the Infinite-Dimensional Hamilton–Jacobi Equation

In this section, we give an infinite-dimensional generalization of the results of Chapter 2 concerning turnpike properties in dynamic optimization problems, as well as of the results of §3.2 concerning the large-time behavior of solutions of the Hamilton–Jacobi equation.

1. Let X be a metric space with metric ρ, and let B_t be the semigroup of operators acting on functions $f\colon X \to A = \mathbb{R}\cup\{+\infty\}$ bounded below according to the formula

$$(B_t f)(x) = \inf_y (b(t,x,y) + f(y)), \tag{3.102}$$

or, in terms of the idempotent operations on A,

$$(B_t f)(x) = \int_X^{\oplus} b(t,x,y)\odot f(y)\,d\mu_{\mathbb{1}}(y),$$

where $t\in\mathbb{R}_+$ (continuous time) or $t\in\mathbb{Z}_+$ (discrete time).

Theorem 3.17 *Assume that the function family $b(t,x,y)$ in Eq.* (3.102) *has the following properties*:

(i) $b(t,x,y)\ge 0\ \forall\, t,x,y$;

(ii) *there exist $\xi_1,\dots,\xi_k\in X$ such that $b(t,x,y)=0$ if and only if $x=y=\xi_j$ for some $j\in\{1,\dots,k\}$*;

(iii) *for any $x\in X$ and $j\in\{1,\dots,k\}$, there exists a t such that $b(t,x,\xi_j)\ne+\infty$ and $b(t,\xi_j,x)\ne+\infty$*;

(iv) *there exists a t_0 such that the functions $b(t_0,\xi_j,x)$ and $b(t_0,x,\xi_j)$ are continuous in x at $x=\xi_j$ for each j*;

(v) *for any neighborhoods* $U_j \subset X$ *of the points* ξ_j *in* X*, we have*

$$\inf\left\{b(t_0,x,y) : (x,y) \notin \bigcup_{j=1}^{k} U_j \times U_j\right\} > 0.$$

Then

(i) *the functions* $b(t,x,\xi_j)$ *and* $b(t,\xi_j,x)$ *have the limits*

$$b_j(x) = \lim_{t\to\infty} b(t,x,\xi_j), \qquad \tilde{b}_j(x) = \lim_{t\to\infty} b(t,\xi_j,x); \tag{3.103}$$

(ii) *the operator family* B_t *is convergent to an operator with a factorizable kernel*; *namely,*

$$\lim_{t\to\infty} b(t,x,y) = \min_j (b_j(x) + \tilde{b}_j(y)). \tag{3.104}$$

Remark 3.8 The statement of the theorem does not include the metric. Actually, the theorem in this form is valid for an arbitrary topological space X. However, to verify the main technical condition (v) (which holds automatically for continuous functions $b(t,x,y)$ on a compact space X), we need some analytical estimates. For example, condition (v) is satisfied if

$$b(t_0,x,y) \geq C \max\Big(\min_j \rho^\alpha(y,\xi_j), \min_j \rho^\alpha(x,\xi_j)\Big) \tag{3.105}$$

with some positive constants c and α.

Remark 3.9 It follows from the theorem that the idempotent operators B_t have the unique eigenvalue $\lambda = 0 = \mathbb{1}$ for all t and that the corresponding eigenspace is finite-dimensional and has the basis $\{b_j(x)\}_{j=1,\dots,k}$.

Proof. (i) It follows from the semigroup property of the operators B_t that

$$b(t+\tau,x,y) = \inf_\eta (b(t,x,\eta) + b(\tau,\eta,y)) \tag{3.106}$$

for any t and τ. Hence,

$$b(t+\tau,x,\xi_j) \leq b(t,x,\xi_j) + b(\tau,\xi_j,\xi_j) = b(t,x,\xi_j)$$

according to (ii). Thus, properties (i) and (ii) imply that the functions $b(t,x,\xi_j)$ and $b(t,\xi_j,x)$ are bounded below and nonincreasing with respect to t, whence assertion (i) of the theorem follows.

(ii) The semigroup property (3.106) implies

$$b(2t,x,y) \leq b(t,x,\xi_j) + b(t,\xi_j,y)$$

for any t, j. Consequently,

$$\overline{\lim_{t\to\infty}}\, b(t,x,y) \leq \min_j (b_j(x) + \tilde{b}_j(y)). \tag{3.107}$$

Furthermore, let $N(t)$ denote the maximum integer in t/t_0. By the semigroup property, we have

$$b(t,x,y) = \inf\{B(t,\eta_1,\dots,\eta_{N(t)}) : \eta_1,\dots,\eta_{N(t)} \in X\},$$

where

$$B(t,\eta_1,\dots,\eta_{N(t)}) = b(t_0,x,\eta_1) + b(t_0,\eta_1,\eta_2) + \cdots + b(t_0,\eta_{N(t)-1},\eta_{N(t)}) + b(t - t_0N(t),\eta_{N(t)},y).$$

Let us say that a tuple $\eta_1,\dots,\eta_{N(t)}$ is ε-*optimal* if

$$|b(t,x,y) - B(t,\eta_1,\dots,\eta_{N(t)})| \le \varepsilon.$$

Consider arbitrary neighborhoods U_j of the points ξ_j. It follows from (3.107) and from condition (3.106) that for each ε-optimal tuple all points $\eta_1,\dots,\eta_{N(t)}$ except for a finite number $K(\varepsilon,\{U_j\})$ (which depends on ε and $\{U_j\}$ but is independent of t) of such points lie in the union $\bigcup_{j=1}^k U_j$. In particular, one can construct functions $T(t) \in [t/3, 2t/3]$ and $\eta(t) \in \bigcup_{j=1}^k U_j$ such that

$$|b(t,x,y) - ((b(T(t),x,\eta(t)) + (b(t-T(t),\eta(t),y))| < \varepsilon. \tag{3.108}$$

Using property (iv), let us choose U_j so that

$$b(t_0,\eta,\xi_j) \le \varepsilon, \qquad b(t_0,\xi_j,\eta) \le \varepsilon$$

for each $\eta \in U_j$.

Using the semigroup property once more, let us write

$$b(T(t),x,\eta(t)) > b(T(t)+t_0,x,\xi_j) - b(t_0,\eta(t),\xi_j),$$
$$b(t-T(t),\eta(t),y) > b(t-T(t)+t_0,\xi_j,y) - b(t_0,\xi_j,\eta(t)).$$

Consequently,

$$\varliminf_{t\to\infty} b(T(t),x,\eta(t)) + b(t-T(t),\eta(t),y) \ge \min_j (b_j(x) + \tilde b_j(y)) - \varepsilon.$$

It follows from this and from Eq. (3.106) that

$$\varliminf_{t\to\infty} b(t,x,y) \ge \min_j (b_j(x) + \tilde b_j(y)) - 2\varepsilon, \tag{3.109}$$

where $\varepsilon > 0$ is arbitrary. Equations (3.107) and (3.109) obviously imply Eq. (3.104). The theorem is proved.

2. Let us now proceed to the infinite-dimensional differential Hamilton–Jacobi equation. Namely, let Φ be an arbitrary locally convex space with a countable base of neighborhoods of zero, so that the topology of Φ can be specified by a translation-invariant metric ρ, and let Φ' be the dual space, that is, the space of continuous linear functionals on Φ. Let $H\colon \Phi\times\Phi' \to \mathbb{R}\cup\{+\infty\}$ be a function convex with respect to the second argument. The equation

$$\frac{\partial S}{\partial t} + H\left(x, \frac{\partial S}{\partial t}(t,x)\right) = 0 \tag{3.110}$$

will be called an infinite-dimensional Hamilton–Jacobi equation, and the function $L\colon \Phi \times \Phi \to \mathbb{R} \cup \{+\infty\}$ given by the formula

$$L(x,v) = \sup_p((p,v) - H(x,p))$$

will be referred to as the *Lagrangian* corresponding to the Hamiltonian H. Let

$$b(t,x,y) = \inf \int_0^t L(q,\dot{q})\, d\tau, \tag{3.111}$$

where the infimum is taken over all continuous piecewise smooth curves $q(\tau)$ such that $q(0) = y$ and $q(t) = x$. The function

$$(R_t S_0)(x) = \inf_y (S_0(y) + b(t,x,y)) \tag{3.112}$$

will be called the generalized solution of the Cauchy problem for Eq. (3.108) with the initial function $S_0(x)$.

Remark 3.10 Under various assumptions about the Hamiltonian H, this definition of a generalized solution can be justified in several ways: one can either construct a sufficient supply of classical solutions following §3.2, or use the results of §3.2 to define generalized solutions of the infinite-dimensional equation (3.110) as limits of solutions of its finite-dimensional approximations, or construct semimodules of generalized functions with a natural action of the differentiation operator following Maslov and Samborskii [**?**], or use the infinite-dimensional version of the vanishing viscosity method. A complete theory that follows the latter approach is developed in the series of papers [**?**, **?**, **?**, **?**, **?**, **?**]. Here we do not dwell on this justification, since in numerous problems (for example, in dynamic optimization problems) formula (3.112) is the primary one (namely, Eq. (3.110) occurs as a corollary of (3.112)). Instead, we study the behavior of the function (3.112) for large t.

Theorem 3.18 *Suppose that the Lagrangian L has the following properties*:

(i) $L(x,v) \geq 0$ *for any* x, $v \in \Phi$;

(ii) *there exist points* $\xi_1, \ldots, \xi_k$ *such that* L *vanishes only at* $(\xi_j, 0)$, $j = 1, \ldots, k$;

(iii) $L(x,v)$ *is bounded in some neighborhoods of* $(\xi_j, 0)$ *and continuous at these points*;

(iv) *there exist neighborhoods* U_j *of the points* ξ_j *in* Φ *such that* $L_k(x,v) \geq c > 0$, *with some constant* c, *for all* $x \in \bigcup_{j=1}^k U_j$ *and all* v *and that*

$$L(x,v) \geq c\rho^\alpha(x,\xi_j) \tag{3.113}$$

for all $x \in U_j$ *with some constants* c *and* α.

Then the operator family R_t *given by Eqs.* (3.111) *and* (3.112) *is a semigroup* (*with continuous time* $t \in \mathbb{R}_+$) *and converges as* $t \to \infty$ *to an operator with factorizable kernel. Thus, the kernel family* (3.111) *satisfies conclusions* (i) *and* (ii) *for Theorem* 3.17.

Proof. First, note that the assumptions of Theorem 3.18 do not imply the main condition (v) in Theorem 3.17, and so Theorem 3.18 is not a special case of Theorem 3.17. Nevertheless, the proof is essentially the same. Namely, conditions (i)–(iii) in Theorem 3.18 obviously imply conditions (i), (ii), and (iv) in Theorem 3.17 for the operators R_t, and consequently, the kernels $b(t,x,y)$ (10) satisfy conclusion (i) in Theorem 3.17 and Eq. (3.107). Moreover, it follows from Eq. (3.113) that each ε-optimal trajectory for $b(t,x,y)$ (that is, a curve that offers the infimum in Eq. (3.111) modulo $O(\varepsilon)$) spends almost all time in $\bigcup_{j=1}^{k} U_j$, where $\{U_j\}$ is an arbitrary collection of neighborhoods of the points ξ_j. It follows from this property that inequality (3.108) is valid, and the remaining part of the proof coincides with that in Theorem 3.17.

3.8. Idempotent Calculus of Variations

In calculus of variations, minimization problems for functionals defined by integrals are reduced to some differential equations (the Euler–Lagrange equations) and vice versa. In particular, numerous studies on linear and nonlinear partial differential equations are based on representing solutions as minima of some integral functionals. In idempotent analysis, integration is replaced by evaluating the supremum (or maximum), and an idempotent analog of calculus of variations must involve a study of differential equations to be satisfied by the solutions of the minimization problem for the functional

$$\begin{aligned} F(u) &= \max_{x\in[a,b]} L(x, u(x), u'(x), \ldots, u^{(k)}(x)) \\ &= \int_{[a,b]}^{\oplus} L(x, u(x), u'(x), \ldots, u^{(k)}(x)), \end{aligned} \tag{3.114}$$

where $[a,b]$ is a closed interval, $L(x,v)$ is a continuously differentiable function, and $v = v_0, v_1, \ldots, v_k$. This study was carried out by M. Bronstein et al. [**?**, **?**, **?**], where the more general situation is considered, in which x is multidimensional and L is a vector function. We present some results of the papers [**?**, **?**, **?**] without proof; details can be found in the original papers.

Consider the function space

$$\begin{aligned} D = D_{B,\lambda} = \{u \in C^k[a,b] : B_j(u(a), \ldots, u^{(k-1)}(a), u(b), \ldots, \\ u^{(k-1)}(b), u^k(a) - \lambda u^k(b)) \geq 0,\ j = 1, \ldots, N\}, \end{aligned}$$

where B_j, $j = 1, \ldots, N$, are arbitrary continuous functions and λ is a positive constant. Suppose that D is not empty and set

$$m = \inf_{u\in D} F(u) \in (-\infty, \infty).$$

Theorem 3.19 ([?]) *Let $\varepsilon > 0$, and let $\dfrac{\partial L}{\partial u^{(k)}}(x, v) \neq 0$ whenever $L(x, v) < m + \varepsilon$. Then the functional (3.114) attains its minimum on D if and only if there exists a sequence $u_i \in D$ bounded in $C^k[a, b]$ and minimizing F, that is, $\lim_{i\to\infty} F(u_i) = m$. Each point of minimum satisfies the equation*

$$L(x, u(x), u'(x), \ldots, u^{(k)}(x)) = m \qquad \forall\, x \in [a, b].$$

This theorem and its vector generalizations permit one to prove existence theorems for boundary value problems for some systems of differential equations. Here we only give two examples of applying these results.

1. Consider the nonhomogeneous Korteweg–de Vries equation

$$u'''_{(x)} + 6u(x)u'(x) = f(x),$$

where f is a T-periodic function with zero mean,

$$\int_0^T f(x)\, dx = 0.$$

This equation has infinitely many periodic solutions, each of which is a point of minimum of the functional (3.114) with $L = u''(x) + 3u^2(x)$ on the set

$$D_\alpha = \{u \in C^2[0, T] : u(0) = u(T),\ u'(0) = u'(T),\ u(0) \leq \alpha\}$$

for some $\alpha = \alpha(\lambda)$.

2. Consider the eigenvalue problem for the Schrödinger operator $-\Delta + V(x)$ on the torus $\mathbb{T}^n$, $V \in C^\infty(\mathbb{T}^n)$. We seek the eigenfunction in the form $u = e^\psi$, which yields the equation

$$-\Delta\psi - \left|\frac{\partial\psi}{\partial x}\right|^2 + V(x) = \lambda$$

for ψ. In terms of the phase function ψ, we can state a new variational principle for the least eigenvalue: the phase function ψ is the minimum on $C^2(\mathbb{T})$ of the functional

$$F(\psi) = \max_{x\in\mathbb{T}} \left(\Delta\psi(x) + \left|\frac{\partial\psi}{\partial x}\right|^2 - V(x)\right),$$

and the least eigenvalue of the Schrödinger operator differs from the minimum of the functional F only in sign.

This variational principle implies the estimate

$$\lambda \leq \frac{1}{\operatorname{mes} \mathbb{T}^n} \int_{\mathbb{T}^n} V(x)\, dx$$

for the least eigenvalue and leads to the following comparison theorem: if $V_1 = q * V_2$, where the asterisk denotes convolution and q is an arbitrary nonnegative function with unit mean, $\int_{\mathbb{T}^n} q(x)\,dx = 1$, then $\lambda_1 \geq \lambda_2$, where λ_j is the least eigenvalue of the Schrödinger operator with potential V_j. The proof, as well as other examples, can be found in the cited papers.

3.9. Limit Theorems for Discrete Idempotent Measures

To clarify the background of the main idea of this section, let us briefly explain how the theory of finite-difference approximations can be used to obtain the central limit theorem very easily.

In the simplest form, this theorem states the following. If $\xi_1, \xi_2, \ldots$ is a sequence of independent random variables that assume two values, 1 and -1, with equal probability, then the distribution of the sequence of random variables $S_n = (\xi_1 + \ldots + \xi_n)/\sqrt{n}$ converges to the normal distribution, that is,

$$\lim_{n\to\infty} P(a \leq S_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx.$$

Let η be a real-valued random variable with continuous distribution density $q(x)$ and independent from $\xi_1, \xi_2, \ldots$. We introduce a discrete approximation of η by random variables η_n, $n \in \mathbb{N}$, that assume the values kh, $k \in \mathbb{Z}$, with probabilities

$$q_k^n = \int_{(k-\frac{1}{2})h}^{(k+\frac{1}{2})h} q(x)\,dx,$$

where $h = 1/\sqrt{n}$. Obviously, the η_n weakly converge to η as $n \to \infty$. Furthermore, the distribution of the sum $\eta_n + S_n$ is the convolution of the distributions of η_n and S_n. Hence, if this sum weakly converges as $n \to \infty$, then the density of the limit distribution is the convolution of $q(x)$ with the limit distribution of S_n.

Let us now construct the simplest finite-difference approximation for the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$$

in the strip region $-\infty < x < \infty$, $t \in [0,1]$ with the initial condition $q(x)$ at $t = 0$. We consider the time increment $\tau = 1/n$ and the space increment $h = 1/\sqrt{n}$ and pose an approximate initial condition $q_n(x)$ by setting $q_n(x) = q_k^n$ for $x \in [(k-1/2)h, (k+1/2)h]$. Then the difference approximation $u_n(m\tau) = \{u_{kh}^{m\tau}\}$ to the solution of the Cauchy problem is determined from the recursion relation

$$u_{kh}^{(m+1)\tau} = \frac{1}{2}(u_{(k-1)h}^{m\tau} + u_{(k+1)h}^{m\tau}), \qquad m \geq 0,$$

and from the initial condition $u_{kh}^0 = q_n(kh)$. Obviously, $u_n(\tau m)$ is the distribution of the random variable $\eta_n + h(\xi_1 + \ldots + \xi_m)$; in particular, $u_n(1)$ is the distribution of $\eta_n + S_n$.

Now it follows from the theory of finite-difference approximations that the sequence u_n converges to the solution of the Cauchy problem for the heat equation with the initial function $q(x)$. Hence, the distribution of $\eta_n + S_n$ converges to the distribution whose density is the convolution of $q(x)$ with the normal law. On the other hand, as was indicated above, the limit distribution of $\eta_n + S_n$ is the distribution whose density is the convolution of $q(x)$ with the limit distribution of S_n. Thus, the limit distribution of S_n is normal.

We see that the theory of linear finite-difference approximations works in two directions. On one hand, one can use linear finite-difference approximations to construct approximate solutions of linear differential equations; on the other hand, the limits (generally, weak) of solutions of linear difference equations can be calculated by solving the corresponding linear differential equation. In this section, we intend to show that there is a similar relationship between the differential Bellman equation ("linear" in the sense of idempotent analysis) and its "linear" finite-difference approximations.

Specifically, mainly following our paper [**?**], we use simple examples to show how information about asymptotic properties of solutions to discrete problems with a large parameter can be obtained from the theory of generalized solutions of the differential Bellman equation, developed in §3.2.

Example 3.6 Consider a simplified version of the computer game "Life," which simulates the evolution and death of a population. Each point of the two-dimensional lattice $\mathbb{Z}^2$ can be in one of the two possible states $\mathbb{1}$ (there is a living creature at the point) or $\mathbb{0}$ (there is not). Suppose that the population evolution obeys the following rule: at each time $k \in \mathbb{N}$ (that is, in the kth generation) a given point is in the state $\mathbb{1}$ if and only if at time $k-1$ (in the $(k-1)$st generation) either its left or its lower neighbor was in the state 1. The same evolution occurs in production scheduling (see §1.2). Suppose that in the zeroth generation the points

$$(m,n) \in \mathbb{Z}^2 : \quad \{m = 0,\, n \geq 0\} \quad \text{or} \quad \{m \geq 0,\, n = 0\},$$

were in the state $\mathbb{1}$ and the other points were in the state $\mathbb{0}$. Obviously, the set of points in the state $\mathbb{1}$ at time k is

$$(\{0 \leq m \leq k\} \cup \{0 \leq n \leq k\}) \cap \{m + n \geq k\};$$

for $k = 3$ this set is shown by small circles in Fig. 1 (a).

Although the process has no limit as $k \to \infty$, for each open set $\Omega \subset \mathbb{R}^2$ we can predict whether the set $k\Omega \cap \mathbb{Z}^2$ contains at least one point in the state $\mathbb{1}$ in the kth generation for large k. It is easy to see that this is so if and only if Ω has a nonempty intersection with the set

$$P = (\{0 \leq x \leq 1\} \cup \{0 \leq y \leq 1\}) \cap \{x + y \geq 1\} \subset \mathbb{R}^2,$$

dashed in Fig. 1 (a).

Fig. 1. (a) (b)

Thus, to each open set $\Omega \subset \mathbb{R}^2$ there corresponds a uniquely determined symbol $\mu(\Omega)$, which is equal to $\mathbb{1}$ or $\mathbb{0}$ depending on whether the kth generation of the population enlives in $k\Omega$ as $k \to \infty$. Let us regard $\mathbb{1}$ and $\mathbb{0}$ as elements of the two-element Boolean algebra B with respect to the usual addition $\oplus$ and multiplication $\odot$ (disjunction and conjunction in propositional calculus). Then $\mu(\Omega)$ is an additive set function with respect to $\oplus$,

$$\mu(\Omega_1 \cup \Omega_2) = \mu(\Omega_1) \oplus \mu(\Omega_2)$$

for any open sets $\Omega_1, \Omega_2 \subset \mathbb{R}^2$, and hence it can be called an *idempotent measure* (with values in B). It is also obvious that the kth generation determines an idempotent measure (an additive B-valued set function) μ_k on $\mathbb{Z}^2$ according to the following rule: for $M \subset \mathbb{Z}^2$, $\mu_k(M) = \mathbb{1}$ if and only if M contains at least one point which is in the state $\mathbb{1}$ in the kth generation. In terms of these idempotent measures, the cited enliving criterion can be stated as the limit equation

$$\lim_{k\to\infty} \mu_k(k\Omega \cap \mathbb{Z}^2) = \mu(\Omega) = \begin{cases} \mathbb{1}, & \Omega \cap P \neq \emptyset, \\ \mathbb{0}, & \Omega \cap P = \emptyset \end{cases} \tag{3.115}$$

for open sets $\Omega \subset \mathbb{R}^2$, that is, as a limit theorem for discrete idempotent measures similar to the limit theorems in probability theory, which establish the convergence of distributions (or probability measures) given by a sequence of discrete random variables to some distribution with continuous density. The enliving criterion (3.115) is trivial. This is not always the case, as shown by the following example.

Example 3.7 Let us consider a different evolution law. Namely, suppose that the point (m, n) is in the state $\mathbb{1}$ in the kth generation if and only if at least one of the points $(m - n, n + 1)$ and $(m - n, n - 1)$ was in the state $\mathbb{1}$ in the $(k - 1)$st generation. Suppose that the initial state of the process (the zeroth generation) is the same as in Example 3.6. Simple considerations show that in the kth generation the points in the state $\mathbb{1}$ lie on the rays issuing from the points

$$\{(m,n) : m = k(-1)^{\varepsilon_1} + (k-1)(-1)^{\varepsilon_2} + \cdots + (-1)^{\varepsilon_k},\\ n = (-1)^{\varepsilon_1} + \cdots + (-1)^{\varepsilon_k},\ \varepsilon_j = 0,1\}$$

with slope $1/k$ or parallel to the axis $\{n = 0\}$.

Let us try to describe the sets $\Omega = \Omega_x \times \Omega_y \subset \mathbb{R}^2$, where Ω_x and Ω_y are open intervals, such that the set $\Omega_k = k^2\Omega_x \times k\Omega_y$ contains points in the state $\mathbb{1}$ in the kth generation as $k \to \infty$. It turns out that this is true if and only if Ω has a nonempty intersection with the set

$$Q = (\{|x-y| \le \tfrac{1}{2}\} \cup \{|y| \le 1\}) \cap \{x \ge \tfrac{1}{4}(y+1)^2 - \tfrac{1}{2}\} \subset \mathbb{R}^2$$

(see Fig. 1 (b), where Q is dashed). The corresponding limit theorem for idempotent discrete measures μ_k determined (as in Example 3.6) by the kth generations has the form

$$\lim_{k\to\infty} \mu_k(k^2\Omega_x \times k\Omega_y \cap \mathbb{Z}^2) = \mu(\Omega) = \begin{cases} \mathbb{1}, & \Omega \cap Q \ne \emptyset, \\ \mathbb{0}, & \Omega \cap Q = \emptyset \end{cases} \tag{3.116}$$

for any open sets $\Omega = \Omega_x \times \Omega_y \subset \mathbb{R}^2$.

Apparently, the limit equation (3.116) can readily be obtained from the explicit expression for the kth generation and for the measure μ_k, by analogy with the derivation of the de Moivre–Laplace limit theorem in probability theory from the Bernoulli distribution by passing to the limit. However, the straightforward evaluation of the limit is rather complicated in this example. We derive Eq. (3.116) from the general theory of the differential Bellman equation by carrying out the following steps.

1) We rewrite the evolution law for the process in question as a discrete Bellman equation, linear in the semimodule of functions $\mathbb{Z}^2 \to B$.

2) We embed B as a subsemiring in the semiring $A = \mathbb{R} \cup \{+\infty\}$ with the operations $a \oplus b = \min(a,b)$ and $a \odot b = a + b$ by identifying the elements $\mathbb{0}, \mathbb{1} \in B$, respectively, with the neutral elements $\mathbb{0} = +\infty$ and $\mathbb{1} = 0$ in A.

3) We consider the auxiliary problem, namely, a process in the semimodule of A-valued functions on $\mathbb{R}^2$ coinciding with the original process at the points of the integral lattice. The discrete Bellman equation satisfied by this process proves to be a finite-difference approximation to some differential Bellman equation.

4) Using Pontryagin's maximum principle and the theory developed in §3.2, we find the generalized solution of the resultant differential Bellman equation in a closed form; this solution is the limit of the discrete process in the idempotent measure, and the set Q is the support of this solution.

Thus, consider the discrete Bellman equation

$$a^{k+1}(M,N) = \min_{v\in V} a^k(M - CN, N - v)$$

on the lattice $M \in \mathbb{Z}^n$, $N \in \mathbb{Z}^n$, where $V \subset \mathbb{Z}^n$ is a finite set and C is an $m \times n$ matrix of integers. Consider the following auxiliary equation for functions of continuous arguments, depending on the parameter $h > 0$:

$$\begin{cases} S_h(t+h, x, y) = \min_{v \in V} S_h(t, x - hCy, y - hv), \\ S_h(0, x, y) = S^0(x, y). \end{cases}$$

For $t = kh$, we obtain

$$S_h(kh, x, y)$$
$$= \min_{v_1, \dots, v_k} S^0\left(x - khCy - (k-1)h^2Cv_1 - \dots - h^2Cv_{k-1}, y - h\sum_{i=1}^{k} v_i\right).$$

Now let

$$kh = t_0, \quad k \to \infty, \quad h = t_0/k \to 0.$$

Then it follows from Theorem 3.3 that the sequence $S_h(t_0, x, y)$ converges in measure of open bounded sets to a generalized solution of the differential Bellman equation

$$\frac{\partial S}{\partial t} + \left\langle Cy, \frac{\partial S}{\partial x} \right\rangle + \max_{v \in V} \left\langle v, \frac{\partial S}{\partial y} \right\rangle = 0. \tag{3.117}$$

This solution has the form

$$S(t_0, x, y) = \inf\{S^0(\xi, \eta) : (\xi, \eta) \in \widetilde{K}(t_0, x, y)\},$$

where $\widetilde{K}(t_0, x, y)$ is the section by the plane $t = t_0$ of the integral funnel of the generalized differential equation

$$\begin{aligned} &\dot{x} = -Cy, && x(0) = x_0, \\ &\dot{y} \in -\operatorname{co} V, && y(0) = y_0, \end{aligned}$$

in which $\operatorname{co} V$ is the convex hull of V, so that

$$\lim_{h \to 0} \inf_{(x,y) \in \Omega} S_n(t_0, x, y) = \inf_{(x,y) \in \Omega} S(t_0, x, y) \tag{3.118}$$

for open bounded sets $\Omega \subset \mathbb{R}^{n+m}$.

Suppose that the initial state of the process a_k is given by a homogeneous function $a^0(M, N)$,

$$a^0(mM, nN) = a^0(M, N) \qquad \forall\, m, n \in \mathbb{N}.$$

Let $S^0(x, y)$ be a function homogeneous in x and y and coinciding with $a^0(x, y)$ at the integral points. Then

$$S_h(kh, kM, kN) = a^k(M, N).$$

It is easy to show that the infimum on the left-hand side in (3.118) can then be taken over the set

$$\{(x, y) \in \Omega \cap h^2\mathbb{Z}^m \times h\mathbb{Z}^n\},$$

which yields the following limit equation for the process a^k:

$$\lim_{k\to\infty} \inf_{(M,N)\in D_k(\Omega)} a^k(M, N) = \inf_{(x,y)\in\Omega} S(1, x, y), \tag{3.119}$$

where $D_k(\Omega)$ is the image of Ω under the action of the operator D_k given by the matrix $\begin{pmatrix} k^2 & 0 \\ 0 & k \end{pmatrix}$.

In particular, the process considered in Example 3.7 obviously satisfies the equation

$$a^{k+1}(m, n) = \min(a^k(m - n, n - 1), a^k(m - n, n + 1)),$$

for which the limit differential equation (3.117) has the form

$$\frac{\partial S}{\partial t} + y\,\frac{\partial S}{\partial x} + \left|\frac{\partial S}{\partial y}\right| = 0.$$

The solution of the Cauchy problem for this equation can be evaluated in a closed form, and Eq. (3.116) follows from Eq. (3.119) (see [**?**] for details).