Nonstandard Linear-Quadratic Decision Making

Tamer Başar, Jun Moon, Jiongmin Yong, and Huanshui Zhang

Abstract—This paper is a compendium to a tutorial session in the 2023 IEEE Conference on Decision and Control (CDC) by the same authors listed here and carrying the same title. The session (as well as this paper) addresses variations around the basic linear-quadratic-Gaussian (LQG) control paradigm, identifies the underlying challenges in such extensions, and discusses some of their resolutions. The paper is organized into four main parts, each one corresponding to a presentation by one of the authors, as the section titles indicate. The first part discusses stochastic control and stochastic games under nonstandard (or nonclassical) information structures, which have elements of signaling and incentive designs, including also variations around the celebrated counterexample in stochastic control due to Witsenhausen. The first part also covers information transmission limitations in control, and rational expectations models arising in economics. The second part considers variations to stochastic LQ Nash and Stackelberg differential games, deriving explicit equilibrium policies expressed by Riccati equations. The third part describes approaches to solving nonstandard LQ control problems to characterize their feedback-type optimal controllers under three different settings, namely (i) irregularity, (ii) delay state equations, and (iii) asymmetric information structures. The last part provides open-loop and closed-loop solvability analyses for LQ control problems, establishing relationships between forward-backward stochastic differential equations, the optimality condition, and (differential/algebraic) Riccati equations.

I. INTRODUCTION

Since the late 1950's, Linear-Quadratic-Gaussian (LQG) theory (also known as stochastic linear-quadratic (LQ) con-

Tamer Başar is with the Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, IL, USA (email: basarl@illinois.edu).

Jun Moon is with the Department of Electrical Engineering, Hanyang University, Seoul, South Korea (email: junmoon@hanyang.ac.kr).

Jiongmin Yong is with the Department of Mathematics, University of Central Florida, Orlando, Fl, USA (email: jiongmin.yong@ucf.edu).

Huanshui Zhang is with the School of Electrical and Automation Engineering, Shandong University of Science and Technology, Qindao, Shandong, P. R. China (email: hszhang@sdu.edu.cn).

Research of Tamer Başar in this topical area was supported in part by ARO MURI under Grant AG285, in part by ARL-DCIST under Grant AK868, and in part by AFOSR under Grant FA9550-19-1-0353.

Research of Jun Moon in this topical area was supported in part by the Korea Institute of Energy Technology Evaluation and Planning and the Ministry of Trade, Industry & Energy of the Republic of Korea (RS-2023-00235742), in part by the National Research Foundation of Korea (NRF) Grant funded by the Ministry of Science and ICT, South Korea (NRF-2017R1A5A1015311, NRF-2021R1A2C2094350), and in part by Institute of Information & communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01373).

Research of Jiongmin Yong in this topical area was supported in part by the NSF Grant DMS-2305475.

Research of Huanshui Zhang in this topical area was supported in part by National Natural Science Foundation of China (Grant: 62250056), in part the Foundation for Innovative Research Groups of National Natural Science Foundation of China (Grant: 61821004), and in part as Major Basic Research of Natural Science Foundation of Shandong Province (Grant: ZR2021ZD14). trol theory) has been the dominating paradigm for feedback control design for linear stochastic systems under noisy state measurements with Gaussian statistics and quadratic performance index [1], [12], [15], [16], [22], [24], [35], [37], [38], [50], [103], [120], [122]. Its salient feature of separation of estimation and control, allowing the optimal controller to be one of the optimal linear-quadratic-regulator (LQR) solution with simply the state replaced by its conditional mean generated by the Kalman filter (or Kalman-Bucy filter), made the optimal design easily implementable. Furthermore, the optimal controller can be implemented by simply solving the associated (differential or difference or algebraic) Riccati equations, instead of solving complex Hamilton-Jacobi-Bellman PDEs (in the continuous time) or Bellman difference equations (in the discrete time).

Various applications of the LQ control framework can be found in economics, finance, engineering, and sciences. For specific applications, we refer the reader to [1], [15], [16], [33], [34], [37], [38], [47], [51], [59], [103], [120], [122], [133], and the references therein. To list a few, the stochastic LQ pollution level control problem was studied in [33], and [133] considered the stochastic LQ mean-variance portfolio optimization problem. The stochastic control problem of electric water heating loads was studied in [51], the stochastic power adjustment problem of wireless communication networks was considered in [47], and the stochastic LQ problem for queuing systems was considered in [59].

Most of the progress on LQ problems has been built on a number of standard conditions (or assumptions), namely 1) "Regularity", that is, the parameters of the state equation and the associated Riccati equation for designing the controller are required to be sufficiently regular, and the parameters in the cost function need to be sign-definite matrices; 2) "Finite dimensions", that is, the system under consideration is described by ordinary differential equations and with no time delays in most cases; and 3) "Symmetry", that is, the feedback information used by different controllers (in a distributed control problem) is consistent. Questions were naturally raised as to whether the earlier-mentioned attractive and appealing features of the LQ theory are retained when such standard conditions or assumptions are violated, or when one takes variations around the basic model.

To capture general variations in LQ control problems, we may consider the cases of "irregular" (singular control), "infinite dimensional systems" (partial differential/time-delay systems) or "asymmetry in information" (decentralized or distributed control). An early paper that brought to the attention of the control community the vulnerability of the LQG-based design to a modest variation in the standard symmetric information structure assumption, or in other words the assumption that the controller has access to full memory on the measurements it receives, was the one by Witsenhausen [111], who produced a counter-example showing that linearity of the optimal controller does not always hold with such variations around the basic LQG model. Specifically, Witsenhausen addressed a 2-stage scalar LQG problem, showing that if the control at stage 2 does not have access to the measurement of the control at stage 1, then the basic linearity property of the standard LQG solution breaks down, and the optimal solution (which exists) is then nonlinear (even though its closed-form optimal structure is still not known).

There are also other nonstandard assumptions or information structures in LQ problems that have been considered in the literature, such as: (i) having distributed control inputs with decentralized information, (ii) having bandwidth constraints (or probabilistic constraints) and/or sporadic failures on the channels that carry state information to the controller(s) and/or controller inputs to the plant (generally known as networked control), (iii) the plant or the channels carrying information being vulnerable to adversarial attacks (that is, worst-case designs, captured via the framework of zero-sum dynamic games), and (iv) having multiple agents with non-aligned objectives interacting over a network (that is, a nonzero-sum differential/dynamic game framework). These classes of LQ problems (hereafter referred to as nonstandard LQ problems) are very complex, with each one requiring a different set of tools for their analyses and synthesis of appropriate optimal solutions (or policies, in a broader framework). Some of these will be discussed in this survey article.

Accordingly, we address in this paper selected nonstandard deterministic and stochastic LQ problems, discussing the underlying challenges and some of their resolutions. The paper is organized into four main parts, along the lines of the presentations of each author in the 2023 CDC tutorial session.¹

The first part (Section II) studies LQ problems under nonstandard (or nonclassical) information structures, where issues of *signaling* and *incentive designs* arise; it also discusses in this context the counterexample (in stochastic control) due to Witsenhausen and some variations around it. This first part also covers *information transmission limitations* in LQ problems caused by communication channels being unreliable or bandwidth-limited. In addition, it covers extensions to mis-aligned objectives (that is, game problems) and variations on rational expectation models in economics. The second part (Section III) considers several generalizations to multi-criteria LQ problems (particularly stochastic LQ Nash and Stackelberg differential games), addressing the "irregularity" condition and deriving the corresponding equilibrium policies expressed by Riccati equations.

The third part (Section IV) describes approaches of solving

LQ control problems to characterize their feedback-type optimal controllers under three different nonstandard settings, namely (i) irregularity condition, (ii) delay state equations, and (iii) asymmetric information structures. The last part (Section V) provides open-loop and closed-loop solvability analyses for LQ control problems, establishing relationships between forward-backward stochastic differential equations, the optimality condition, and (differential/algebraic) Riccati equations. The paper ends with some concluding remarks, and an extensive list of references.

II. VARIATIONS AROUND THE STANDARD LQG MODEL (BY TAMER BAŞAR)

As we have mentioned in the Introduction, and as is well known in the broad community that deals with decisionmaking under uncertainty (not only control), the LQG model (linear systems dynamics driven by control and Gaussian independent increment noise process, where the control has access to state information over a linear Gaussian channel, and with a performance index that assigns quadratic positive costs to the state and the control over the decision horizon) admits a very appealing and easily implementable optimal solution, where the optimal controller is the one of the LOR (which is deterministic version of the LOG) with only the state replaced by its least-squares estimate, generated by the Kalman filter which is a dynamic system of the same dimension as the system state, driven by the associated innovation process. This is true for both discrete- and continuous-time formulations, and finite as well as infinite decision horizons, with the implementation requiring the solutions of two (uncoupled) Riccati equations (discrete-time or continuous-time or algebraic, as appropriate). Stripping away the details, one important property of the optimal solution is that at any point in time the optimal controller is a linear (affine if random quantities have nonzero means) function of all the past measurements made, with each past measurement appropriately weighted; in other words, the solution requires access to full memory. Of course, the Kalman filter conveniently prevents the "storage" from growing with time, by generating the least-squares estimate through a dynamical system of fixed dimension.

In this section we discuss several variations around that basic model, which relax the modeling assumptions made in different directions, such as bringing restrictions on memory, allowing measurement channels to fail sporadically, and the decision variables being controlled by different agents with misaligned interests (that is multi-criteria game formulations, including incentive design problems)–all in the linear-quadratic setting. To convey the message(s) in simplest possible settings without much clutter of notation, we stay with discrete-time formulation, with two decision makers, scalar quantities, but of course with dynamic information, which still provides a rich landscape.

A. Restricted Memory

We start with the following two-stage stochastic control or equivalently two-agent dynamic stochastic team problem,

¹In the paper, by necessity different notations and terminologies will be used in different (sub)sections, but this difference will have minimum to no impact on the delivery of the main contents.

where all quantities are scalar: A Gaussian random variable, x, with mean zero and variance σ_x^2 is to be transformed into another random variable, $u_0 = \gamma_0(x)$, which is transmitted over a Gaussian channel, $y = u_0 + w$, with zero-mean additive Gaussian noise w of variance σ_w^2 , the output, y, of which is to be further transformed into another random variable, $u_1 = \gamma_1(y)$. The objective is to choose the transformations γ_0 and γ_1 in such a way that a performance index, $Q(x, u_0, u_1)$, quadratic in x, u_0 , and u_1 , is minimized in the average sense. That is, we seek the pair $\gamma^* := (\gamma_0^*, \gamma_1^*)$ such that

where

$$J(\gamma) = \mathbb{E}\left[Q(x, \gamma_0(x), \gamma_1(y))\right]$$
(II.2)

with expectation, $\mathbb{E}[\cdot]$, taken over the statistics of x and w, which are assumed to be independent. Furthermore, the minimization is over the space of all Borel-measurable maps, that is both policies (decision rules) γ_0 and γ_1 are taken to be Borel-measurable maps of the real line onto itself.

 $J(\gamma^*) = \min_{\gamma} J(\gamma) =: J^*,$

This is a stochastic decision problem with what is known as *nonclassical information*, because the information to be used by the decision rule, γ_1 , of the second agent depends on the action, u_0 , of the first agent (and thereby on the decision rule of the first agent), but the second agent does not have access to the information of the first agent (that is, x). If we view it as a single-agent problem where the agent acts twice, then it is one where the agent is memoryless, that is it does not remember what it had observed at the earlier stage. As such, these problems for which a systematic solution process does not exist, one of the main reasons being that due to loss in memory, a sequential decomposition is not possible.

Note that if the second agent also had access to x, then ywould provide useless information, and we would essentially have a deterministic problem, admitting a linear optimal solution (linear in x) for both agents (assuming that Q is strictly convex in the pair (u_0, u_1) . If the first agent did not have perfect access to x, but observed it over a noisy channel, such as $y_0 = x + w_0$, where w_0 is another zeromean Gaussian random variable, independent of the others, and if y_0 was also accessible to the second agent (that is, $u_1 = \gamma_1(y_0, y), u_0 = \gamma_0(y_0)$), then this fall within the setting of LQG, and hence optimum γ_1 and γ_0 would be linear in their arguments (again assuming the earlier strict convexity condition). But with the memory restriction as above, linearity of the optimal policies is no longer a foregone conclusion, and as we discuss below it very much depends on the structure of Q.

We now consider different instances of this class of problems, corresponding to different choices of the performance index Q, some of which admit explicit, relatively simpler solutions while some others do not. Hence the message will be that it is not only the nonclassical nature of the information structure but also the structure of the performance index that contributes to the difficulty in solving these problems.

Let us start with the most general quadratic structure for Q in terms of the 3 scalar variables (x, u_0, u_1) ;

$$Q(x, u_0, u_1) = c_0(u_0)^2 + c_1(u_1)^2 + au_0u_1 + b_0xu_0 + b_1xu_1,$$

where a, b_0, b_1, c_0, c_1 are scalar parameters. Note that we have not included a quadratic term in x since it is not relevant to (enter into) the optimization problem faced by the two agents. Our focus here in three different sets of choices of these parameters, with each set leading to an intrinsically different optimization problems with totally different features and requiring totally different solution techniques:

Set 1 (WIT):

(II.1)

 ${c_1 = 1, c_0 = 1 + k_0, b_0 = -2k_0, b_1 = 0, a = -2},$ where $k_0 > 0$ is some parameter.² This leads to the following special structure for Q:

$$Q_W(x, u_0, u_1) = k_0(u_0 - x)^2 + (u_1 - u_0)^2.$$
 (II.3)

Set 2 (GTC):

 $\{c_1 = 1, c_0 = k_0, b_1 = -2, a = -0\}$, where $k_0 > 0$ is some parameter.³ This leads to the following special structure for Q:

$$Q_{\text{GTC}}(x, u_0, u_1) = k_0(u_0)^2 + (u_1 - x)^2 + b_0 u_0 x$$
, (II.4)

Set 3 (ZSG):

 ${c_1 = 1, c_0 = 1 + k_0, b_0 = -2k_0, b_1 = 0, a = -2},$ where $k_0 < 0$ is some parameter.⁴ This leads to the following special structure for Q:

$$Q_W(x, u_0, u_1) = k_0(u_0 - x)^2 + (u_1 - u_0)^2, \ k_0 < 0.$$
 (II.5)

Note that since $k_0 < 0$, here the roles of the agents (players) have changed. Now u_0 is the maximizing variable, while u_1 is still the minimizing variable. This makes the problem a *zero-sum game* with the natural solution concept being one of *saddle-point equilibrium*. Hence, (II.1) is replaced here with

$$J(\gamma_0, \gamma_1^*) \le J(\gamma_0^*, \gamma_1^*) \le J(\gamma_0^*, \gamma_1),$$
(II.6)

where $J(\gamma_0^*,\gamma_1^*)=:J^*$ is the saddle-point value.

We now discuss the solutions to these three different classes of problems, all variations to the general two-agent LQ decision problem formulated at the beginning of this subsection, in the same order as they are listed.

1-WIT:

This is essentially the Witsenhausen counterexample (WC) mentioned earlier in the paper, which is a 2-stage LQG control problem with no memory for the controller [111]. It is

²Here $c_1 = 1$ is just normalization. Our discussion on the nature of the solution to WIT will actually apply to a much larger set, where Q is strictly convex in the pair (u_0, u_1) and $a \neq 0$, that is $c_1 > 0$, $4c_1c_0 > a^2 > 0$.

³Here again $c_1 = 1$ is just normalization. Our discussion on the nature of the solution to GTC will apply to a much larger set where the only requirement is that a = 0 and Q is strictly convex in the pair (u_0, u_1) , that is both c_1 and c_0 be positive.

⁴Here again $c_1 = 1$ is just normalization. Our discussion on the nature of the solution to ZSG will again apply to a much larger set, where $c_1 > 0$ and $a \neq 0$. Note that it is not required that $c_0 > 0$.

an invertible linear transformation that takes WC to the form given here, and the two formulations are equivalent as far as the optimal solution goes, as shown in [20] and [13]. Note that the interpretation in this reformulation is that the first agent (controlling u_0) wants to stay as close to x as possible, while the second agent (controlling u_1) wants to stay as close to the action of the first agent, u_0 , as possible. Witsenhausen has shown in [111] that the optimal solution to this problem exists, but the optimum decision rules are not linear. For the latter, he has shown that there exist nonlinear policies which outperform the best linear ones. A class of such nonlinear policies introduced by Witsenhausen, and further improved upon by Bansal and Başar [20] is

$$\begin{aligned} u_0 &= \gamma_o(x) = \epsilon \operatorname{sgn}(x) + \lambda x \,, \\ u_1 &= \gamma_1(y) = \mathbb{E}[\epsilon \operatorname{sgn}(x) + \lambda x | y], \end{aligned}$$

where ϵ and λ are parameters to be optimized over (in [111] the values are picked as $\lambda = 0$ and $\epsilon = \sigma_x$, and some asymptotics are studied). Clearly, if $\epsilon = 0$, this class of decision rules will be linear, since $\mathbb{E}[\lambda x|y]$ will be linear for each λ , however when $\epsilon \neq 0$, the decision rules at both stages will be nonlinear. For further discussion on this problem on progress made since 1987, see [127].

2-GTC:

The second special structure introduced above arises in communications- (with $b_0 = 0$)-optimum transmission of a (Gaussian) random variable from a source to a destination over a noisy channel, for least squares estimation of that random variable at the destination, which involves optimum designs of an encoder (u_0) and decoder (u_1) . More specifically, here the second agent's objective is to estimate the random variable x in the minimum mean square (MMS) sense, using a measurement that is transmitted over a Gaussian channel where the input to the channel is *shaped* by the first agent who has access to x and has a soft constraint $(k_0 \mathbb{E}[(u_0)^2])$ on its action. The version of this problem, where the soft constraint is replaced by a hard power constraint, $\mathbb{E}[(u_0)^2] \leq$ k, is known as the Gaussian Test Channel (GTC), and in this context γ_0 is the *encoder* and γ_1 the *decoder*, whose optimal choice is clearly the conditional mean of x given y, that is $\mathbb{E}[x|y]$. The best encoder for the GTC can be shown to be linear (a scaled version of the source output, x), which in turn leads to a linear optimal decoder. The approach here (which is in fact the only one known to apply to this problem) is to obtain bounds on the performance using an inequality from information theory involving channel capacity and rate distortion function, and then showing that the bound can be achieved using linear policies. This can be applied to the generalized version formulated here (with $b_0 \neq 0$), leading to the following solution:

$$\gamma_0^*(x) = -\operatorname{sgn}(b_0) \frac{\sqrt{\alpha^*}}{\sigma_x} x, \qquad (\text{II.7})$$

$$\gamma_1^*(y) = \mathbb{E}[x|y] = -\frac{\operatorname{sgn}(b_0)\sigma_x\sqrt{\alpha^*}}{\alpha^* + \sigma_w^2}y, \quad \text{(II.8)}$$

where α^* is the unique positive solution of the polynomial equation

$$[2k_0\sqrt{\alpha} - |b_0|\sigma_x] [\alpha + \sigma_x^2]^2 = 2\sigma_w^2 \sigma_x^2 \sqrt{\alpha}.$$
 (II.9)

Details can be found in [20], [13], and [127]. Extensions to the case of multiple channels between the two agents can be found in [19].

Remark II.1 The main difference between the two problems WIT and GTC is that Q in the former has a product term between the decision variables of the two agents while in the latter it does not. Hence, it is not only the dynamic (nonclassical) nature of the information structure but also the structure of the performance index (cost function) that determines whether linear policies are optimal in LQG multistage decision problems.

The third special structure is similar to WIT but with the roles of the two agents being conflicting, which is captured by k_0 being negative; as indicated earlier this constitutes a zero-sum game with u_0 now maximizing. First note that if γ_0 is linear in x, say $\gamma_0(x) = \alpha x$ for some parameter α , then the γ_1 that minimizes J is also linear (and unique), being the conditional mean of αx given y. Hence,

$$\gamma_0(x) = \alpha x \quad \Rightarrow \quad \gamma_1(y) = \frac{\alpha^2 \sigma_x^2}{\alpha^2 \sigma_x^2 + \sigma_w^2} y$$

Conversely, if γ_1 is linear in y, say $\gamma_1(y) = \lambda y$ for some parameter λ , then provided that

$$k_0 < -(\lambda - 1)^2,$$
 (II.10)

which makes Q_G strictly concave in u_0 ,⁵ the γ_0 that maximizes J is also linear (and unique). Hence,

$$\gamma_1(y) = \lambda y \quad \Rightarrow \quad \gamma_0(x) = -\frac{k_0}{k_0 - (\lambda - 1)^2} x$$

For these policies to constitute a saddle point, we have to find a pair (α, λ) which simultaneously solve

$$\lambda = \frac{\alpha^2 \sigma_x^2}{\alpha^2 \sigma_x^2 + \sigma_w^2}$$
$$\alpha = -\frac{k_0}{k_0 + (\lambda - 1)^2},$$

while satisfying the constraint (II.10).

Following these lines, it has been shown in [13] that the game has a unique saddle-point solution as long as $k_0 < 0$, and the saddle-point policies are linear:

$$\gamma_0^*(x) = -\frac{k_0}{k_0 + (\lambda^* - 1)^2} x \,, \quad \gamma_1^*(y) = \lambda^* y.$$

where λ^* is the unique solution of the polynomial equation $f(\lambda) = 0$ in the interval $(\max(0, 1 - \sqrt{-k_0}), 1)$, where

$$f(\lambda) := (\sigma_w^2 / \sigma_x^2) \lambda \left[k_0 - (1 - \lambda)^2 \right]^2 + k_0^2 (1 - \lambda) = 0.$$

Now uniqueness of the linear saddle-point solution in the

⁵If the condition (II.10) is not satisfied, then for the given γ_1 , the player who chooses u_0 can make the value of J arbitrarily large.

class of general Borel-measurable policies (and not only linear) follows from the *ordered interchangeability* property [16] of multiple saddle points, since the optimum response of each player to an announced linear policy of the other player is unique, as indicated above.

Extension to noisy access to x:

If u_0 has access to noise corrupted version of x, that is $u_0 = \gamma_0(z), z = x + v$, where v is a zero mean Gaussian random variable, independent of x and w, then the difficulty in obtaining the optimum solution to WIT is naturally still there. For GTC and ZSG, on the other hand, optimum solutions exist and the corresponding policies are still linear. Details can be found in [13]. An important point to be made here is that in these two classes of problems (GTC and ZSG) certainty equivalence does not hold, that is the solution is not of the type where one first solves the deterministic version of the problem and then replaces the random variables with their conditional mean values at the solution point. On the other hand, one would normally expect certainty equivalence to hold if the information structure is of the classical type, that is (in this case) the agent acting at the second stage has access to not only his private information ybut also z. This is indeed the case with team problems (which then become standard LQG stochastic control problems), but not necessarily for stochastic zero-sum games which feature many pitfalls; for details we refer to [4] and [3].

Two extensions to GTC:

For the second setting, GTC, discussed in this subsection, one important extension would be to have the channel between the two agents (encoder and decoder) to be susceptible to adversarial action under some second moment constraint, that is in addition to the Gaussian channel noise there is also a *jamming* input to the channel, which could be correlated with x (or z in the extended version). This then becomes a 3-player zero-sum game, where the original two agents (encoder and decoder) form a team (with still nonclassical information) who minimize the objective function, whereas the jammer (as the third agent) maximizes the same under a hard constraint or a soft quadratic constraint attached to the objective function. The solution concept is as in ZSG, that of saddle point with the team of encoder-decoder playing against the jammer. It has been shown in [5] that the saddlepoint solution for this quadratic decision problem is still linear for the encoder-decoder pair, while for the jammer the maximizing policy is to inject a zero-mean Gaussian noise, possibly correlated with x (or z).

A second extension to GTC involves a multi-stage formulation which entails a joint design of the control policy and the measurement process. Consider the scalar discrete-time plant

$$x_{n+1} = \rho_n x_n + u_n + v_n$$
, $n = 0, 1, \dots$

along with the scalar measurement

$$y_n = h_n + w_n$$
, $n = 0, 1, \dots$,

where $\{v_n\}$ and $\{w_n\}$ are i.i.d. Gaussian random variables,

with zero mean and independent of each other as well as of the Gaussian initial state x_0 . The variable u_n is the control, allowed to depend on the present and past values of y, and h_n is another decision variable (the sensor structure), which has to be designed as a function of the current value of the state, x_n , and possibly also of the past values of y, and this design has to be picked optimally, along with the control, so as to minimize the expected value of a stageadditive quadratic cost function. This is a dynamic decision problem that features nonclassical information because u_n and h_n can be seen as the actions of two agents with the decision of one affecting the information of the other, who however do not share information. Employing the GTC result sequentially, as well as sequential rate distortion theory, this nonclassical stochastic control problem can be shown to admit a linear optimal solution (for both u_n and h_n) [21]. Its continuous-time version (again scalar) also admits a linear optimal solution [14], where now the continuous-time Gaussian test channel with feedback is employed. None of these results admit easy extensions to multivariable systems, where optimum solutions (if they exist) will in general not be linear.

B. Rational Expectation Models

Rational expectation models feature special types of state dynamics that are driven by expectations of the future based on current information or by control that is designed as a result of optimizing a cost function that has multi-step future value of the state. These dynamic models are called "forward looking", because the future behavior depends explicitly on the expectations the agents have on the future itself (as it happens in the stock market); and they are also called "rational expectations models", because the expectations on the future outcomes are (or should be) formed on some rational basis. For a background on such models, and for details of the results given in this subsection, we refer the reader to [10]. Perhaps the simplest such model is described by the scalar difference equation:⁶

$$y_t = ay_{t-1} + bE_{t-1}y_{t-1} + \epsilon_t$$
, (II.11)

where a and $b \neq 0$ are constant parameters, $\{e_t\}$ is a sequence of independent zero-mean random variables with finite variance, and $E_{t-1}y_{t+1} := \mathbb{E}[y_{t+1}|\eta_t]$ is the conditional expectation of y_{t+1} based on some information, η_t , available at time t. The subscript t-1 is used to capture the assumption that this information η_t is based on the past is based on the past values of the relevant state (of, say, of the economy), that is $\{y_{t-1}, y_{t-2}, \ldots\} =: y^{t-1}$. A common assumption is to let $\eta_t = y^{t-1}$; but other formulations are also possible, such as $\eta_t = z^{t-1}$, where z_t denotes some "noisy" measurement on y_t :

$$z_t = y_t + \xi_t \,, \tag{II.12}$$

⁶We use here notation more in line with the literature on rational expectations, such as the one adopted in [10], with state denoted by y instead of the common control usage of x, but this should not create any confusion as the material here is self-contained.

with $\{\xi_t\}$ being another sequence of independent, zero-mean random variables with finite variance.

The basic question addressed particularly in the economics literature, rephrased in the above context, is whether there exists a (unique) stochastic process $\{y_t\}$ that satisfies (II.11) for all t of interest. It actually turns out that the solution is actually not unique, which motivated a control-theoretic approach to the problem, as introduced in [10]. We replace (II.11) with the *controlled* state equation:

$$y_t = ay_{t-1} + bv_t + \epsilon_t, \tag{II.13}$$

where $\{v_t\}$ represents an aggregate decision variable, chosen under the information restriction that $v_t = \gamma_t(\eta_t)$, for some Borel-measurable function γ_t -the policy variable. Connecting this with the earlier formulation, a rational choice for γ_t would be to pick it so that v_t is as close to $\mathbb{E}[y_{t+1}I\eta_t]$ as possible. A cost function that would capture this over a horizon (0, T] is

$$J(\gamma) = \sum_{t=0}^{T} \mathbb{E}[(\gamma_t(\eta_t) - y_{t+1})^2] \rho^t, \qquad \text{(II.14)}$$

where $\rho \in (0,1)$ is a discount parameter. The gaol here is to minimize $J(\gamma)$ by properly choosing $\gamma := \{\gamma_0, \ldots, \gamma_T\},\$ where the time horizon could also be infinite. Let us further take the underlying statistics of the random variables to be Gaussian. This is clearly a dynamic policy optimization problem with LQG structure, but not of the standard type. Still, its optimum solution can be shown to exist (whenever ab < 1/4), is unique, and linear in the information available to the controller. In the perfect state measurement case, the optimum controller is linear on the most recently available value of y and the most recently applied control. In the noisy measurement case, the most recently available value of y is replaced by its conditional mean. The finite-horizon optimal controller has a well-defined limit as $T \to \infty$, which is the stationary, stabilizing optimal controller for the infinitehorizon problem (valid for both perfect and imperfect state measurements). These results also admit comparable extensions to the more general case where both the state process yand control v are vector-valued; details can be found in [10]. There are also extensions (and substantial ones) to forwardlooking models where the original uncontrolled dynamics (II.11) has an additional control input that aims at driving the dynamics to a target value using an appropriate quadratic cost function [9].

A second type of model with conditional expectations in the cost:

A second type of variation around the LQG, of the nonstandard type, which also has explicit dependence on conditional mean (as above) has been addressed in [8]. The problem involves active learning, and is motivated by a macro-economics model of credibility and monetary policy, incorporating asymmetric information between the private sector and the monetary authority. The former is a passive player who simply forms conditional (rational) expectations of the current inflation rate, which constitutes the surprise component of the policy maker's (the monetary authority) objective function. The policy maker attempts to maximize the objective function by choosing a control policy which also affects the information carried to the passive player whose rational expectations in turn influence the performance of that policy. More precisely (deviating a bit from the notation of [8], to be somewhat consistent with the above, particularly by turning the original maximization problem into a minimization one), the underlying multistage stochastic decision-making (control) is formulated as follows (again staying with the scalar version): Minimize, over $\gamma := {\gamma_0, \ldots, \gamma_T}$, a Borel-measurable function, the objective functional

$$J(\gamma) = \sum_{t=0}^{T} \mathbb{E}[\frac{1}{2}(u_t)^2 + x_t(\mathbb{E}[u_t|\eta_t] - u_t)]\rho^t, \quad \text{(II.15)}$$

where $\rho \in (0,1)$ is again the discount factor, $u_t = \gamma_t(\eta_t, x_t), t = 0, 1, \ldots$ is the control variable, $\eta_t = y^{t-1}$: $\{y_0, \ldots, y_{t-1}\}, \{x_t\}$ is the state process generated by

$$x_{t+1} = ax_t + c_t + \epsilon_t , t = 0, 1, \dots,$$
 (II.16)

and $\{y_t\}$ is the measurement process generated by

$$y_t = u_t + \xi_t \,. \tag{II.17}$$

The random variables x_0 , $\{\epsilon_t\}$, $\{\xi_t\}$ are Gaussian, independent, with x_0 having nonzero mean, and others having zero mean. Finally, c_t is a nonzero scalar.

We note that the control $\{u_t\}$ enters the problem not through the state equation (II.16), but through the message (or measurement) process (II.17), and the cost function (II.15) to be minimized. The presence of the conditional expectation term in (II.15) makes this a nonstandard stochastic dynamic optimization problem of the first type considered in this subsection; furthermore, the problem is what is known as *non-neutral* since the choice of $\{u_t\}$ has a direct effect on the content of the information carried by the measurement process (II.17) regarding the state process $\{x_t\}$, as it has been the cases of WIT and GTC in Subsection II-A.

No standard approach of stochastic control can be applied to this nonstandard dynamic stochastic optimization problem to obtain its solution, and even to prove its existence. An indirect approach has been developed in [8], which relates the original single-person optimization problem to a sequence of nested zero-sum games, for which existence of a unique saddle-point solution has been shown (under some specific conditions), with one of the policies in the saddle-point pair being the policy that minimizes $J(\gamma)$. One of its unique aspects of this approach is the demonstration of the utility of the powerful machinery of saddle-point equilibria even in problems which are neither formulated as, nor can directly be converted to, zero-sum games. Uniqueness here follows from the ordered interchangeability of multiple saddle points (as in the case of ZSG problem of Subsection II-A, since it turns out that proposing linear solutions for this game lead to unique responses. Finally, as already hinted just now, the minimizing solution to (II.15) is linear, and actually in the form

$$u_t = \gamma_t(x_t, \eta_t) = L_t(x_t - \mathbb{E}[x_t|\eta_t]), \ t = 0, 1, \dots,$$

where $\{L_t\}$ can be computed off-line. For the infinitehorizon problem (that is, as $T \to \infty$) this sequence converges, providing the stationary minimizing solution to $J(\gamma)$ with $T = \infty$. Details can be found in [8].

C. LQG Control and Zero-Sum Games with Channel Constraints

We consider in this subsection two other types of variations around the LQG model, brought about by restrictions on the communication or information transmission channels. One of these (which most of the discussion below will pertain to, and in discrete time) is probabilistic failure of measurement channels, governed by Bernoulli processes. The second one captures a different type of restriction on the transmission channels, which is a constraint on the bandwidth, necessitating quantization schemes to be developed.

Now, for the former type of restriction, we will in fact consider the more general formulation of a *zero-sum dynamical game* where the linear evolution of the state is driven by the controls of two players (agents) with totally conflicting objectives. More precisely, state's evolution is described by

$$x_{t+1} = Ax_t + Bu_t + Dv_t + Fw_t, \ t = 0, 1, \dots, \quad \text{(II.18)}$$

and the measurement equation is

$$y_t = \beta_t (Hx_t + Gw_t), \ t = 0, 1, \dots,$$
 (II.19)

where x_0 is a zero-mean Gaussian random vector; $\{w_t\}$ is a zero-mean Gaussian process, independent across time and of x_0 ; and $\{\beta_t\}$ is a Bernoulli process, independent across time and of x_0 and $\{w_t\}$, with Probability $(\beta_t = 0) = p$, $\forall t$. This essentially means that the channel that carries information on the state to the players, which is noisy, fails with equal probability p at each stage, and these failures are statistically independent. A different expression for (II.19) which essentially captures the same would be

$$y_t = \beta_t H x_t + G w_t, \ t = 0, 1, \dots,$$
 (II.20)

where what fails is the sensor that carries the state information to the channel and not the channel itself. In this case, when $\beta_t = 0$, then this means that the channel only carries pure noise, which of course is of no use to the controllers.

Now, if the players are aware of the failure of the channel or of the sensor when it happens (which we assume to be the case), then the control policies for the players are

$$u_t = \gamma_t(y^t, \beta^t), \quad v_t = \mu_t(y^t, \beta^t), \quad t = 0, 1, \dots,$$
 (II.21)

where $\{\gamma_t\}$ and $\{\mu_t\}$ are measurable functions of appropriate dimensions; let us denote the spaces where they belong respectively by Γ and M.

The quadratic performance index for the players in this

game is taken as

$$J(\gamma,\mu) = \mathbb{E}\left\{\sum_{t=0}^{T-1} [|x_{t+1}|_Q^2 + \lambda |u_t|^2 - |v_t|^2]dt\right\} \quad (\text{II.22})$$

with $u = \gamma(\cdot), v = \mu(\cdot)$, where the expectation is over the statistics of x_0 , $\{w_t\}$ and $\{\beta_t\}$,. Further, $|x|_Q^2 := x^T Q x$, Q is non-negative definite matrices, and $\lambda > 0$ is a scalar parameter. Note that any objective function with nonuniform positive weights on u and v can be brought into the form above by a simple rescaling and re-orientation of u and v and a corresponding transformation applied to B and D, and hence the structure in (II.22) does not entail any loss of generality as a quadratic performance index.

The problem of interest is to find conditions for existence and characterization of saddle-point strategies, that is $(\gamma^* \in \Gamma, \mu^* \in \mathcal{M})$ such that

$$J(\gamma^*, \mu) \le J(\gamma^*, \mu^*) \le J(\gamma, \mu^*), \quad \forall \ \gamma \in \Gamma, \mu \in \mathcal{M}.$$
(II.23)

We do not provide here the complete solution to this problem, but just a few comments on several special cases;

- (i) By taking F and G to be orthogonal to each other, one captures the special case (more common in LQG) where the system and measurement noises are independent.
- (ii) If D = 0, then the maximizing player is no longer present in the game, and this then becomes a stochastic control problem with intermittently failing measurement channel; for the solution to this problem, see [48], where it is shown that the optimum controller is of the certainty-equivalent type, with the state in the LQR solution being replaced by its best estimate given the channel failures; the paper also solves the problem where transmission of control signals to the plant is also done over channels that intermittently (and independently) fail, in which case an adjustment to the Riccati equation of LQR needs to be made.
- (iii) If further (that is in addition to D = 0) p = 0, that is if the channel never fails, then what we have is the standard LQG setting.
- (iv) If p = 0 but $D \neq 0$, then we have a standard stochastic zero-sum dynamic game of the LQG type, which has been discussed in [4], (see also [3]) showing that the appealing certainty equivalence and separation properties of LQG do not hold in the entire parameter space.
- (v) Finally, if H = I and G = 0, we have the case where the players have intermittent access to perfect state measurements, which has also been discussed in [3]; it has been shown within the context of a 2-stage scalar problem, existence of a saddle point depends on the value of p (it should be less than a certain threshold, that is the failure probability should not be high) as well as on λ (it should also be relatively small, that is not a heavy toll on the effort level of the minimizer).

The second type of variation we mentioned in the opening paragraph of this subsection entails bandwidth constraints, necessitating appropriate quantization of signals before they are are inputted the transmission channels which could be links (in a control system) from sensors to controllers as well as from controllers to the plants. The main question that is raised (and is relevant) in this context is the design of the best (optimum) quantization schemes and the corresponding policies so that given the bandwidth constraints optimum system performance is obtained (for, for example, an LOG system). The turn of the millennium has witnessed rapid growth in the number of papers that have contributed to this topical area; these developments, until circa 2013, have been discussed in a comprehensive way in the book by Başar and Yüksel [127], One of the results from this period addresses rate requirements for state estimation in discrete-time linear time-invariant (LTI) systems where the controller and the plant are connected via a noiseless channel with limited capacity [125], establishing (using information-theoretic arguments) the existence of optimal variable-length and fixedlength quantizers and construction of such optimal quantizers under three different stability criteria (namely, monotonic boundedness of entropy, asymptotic stability of distortion, and support size stability). It turns out that the uniform quantizer is, in addition to being simple, quite efficient in linear control systems. Another set of results, this time involving continuous-time stochastic LTI systems (driven by Brownian motion) and communication takes place over noisy memoryless discrete- or continuous-alphabet channels, featuring noise in both the forward channel (connecting sensors to the controller) and the reverse channel (connecting the controller to the plant) [126]. One of the main messages is that for stability of the closed-loop system, it is necessary that the entire control space and the state space be encoded, and that the reverse channel be at least as reliable as the forward channel.

D. Multi-criteria Variations

Another variation from the LQG framework entails going from single agent (single controller) formulation to multiple agents (controllers, players) each having a different objective to optimize, with each objective function depending not only on self decisions/actions but also on the decisions of (at least a subset of) other agents. Within the natural noncooperative framework, this brings us to the setting of nonzero-sum dynamic games, where either Nash equilibrium or Stackelberg equilibrium is adopted as a solution concept depending on whether the mode of decision making is symmetric (Nash equilibrium) or asymmetric/hierarchical (Stackelberg equilibrium) [16]. Even within the linear-quadratic-Gaussian framework such a departure from single criterion to multiple criteria brings along many challenges particularly under asymmetric information among the players, since each player would be second-guessing others, trying to decipher the information that others have from their actions, which could be useful in improving his performance-a process that leads to an infinite recursion-still an active area of research. A special case of such problems where there are only two players, and with totally conflicting goals (that is the setting

of zero-sum games) does not exhibit all these challenges because of some appealing properties of saddle-point equilibrium (SPE), as we have seen earlier in this section. For games that do not have the zero-sum structure, even with two players, these appealing properties of SPE disappear when it comes to Nash or Stackelberg equilibrium. Even multiplayer problems with a single objective function would lead to a nonzero-sum game when the players *do not see the world the same way* and have different subjective probabilistic descriptions of the underlying random variables; see [7] for extensive discussion and analyses of games with multiple probabilistic models.

In this subsection we focus on one class of multi-criteria hierarchical (Stackelberg) stochastic decision problems with dynamic information, that arises in incentive designs, which presents its own challenges. Perhaps the simplest such problem with two agents that captures the intricacies of dynamic Stackelberg games (also known in this context as incentive design problem) is the following: There are two agents (players), leader (L) and follower (F), with action variables, u_L and u_F , respectively. Consider three second-order random variables, (x, y, z), with a known joint distribution (such as Gaussian, known to both agents). L and F have different objective functionals, $Q_L(x; u_L, u_F)$ and $Q_F(x; u_L, u_F)$, respectively. The follower has access to the realized value of z and the leader has access to (y, z, u_F) , that is the information structure of the game is $\eta_L = \{y, z, u_F\}$ for L and $\eta_F = \{z\}$ for F, so that $u_L = \gamma_L(\eta_L)$ and $u_F = \gamma_F(\eta_F)$, where γ_L and γ_F are policies for L and F, picked as general measurable maps, belonging to appropriately constructed policy spaces, Γ_L and Γ_F , respectively. Let $J_L(\gamma_L, \gamma_F)$ and $J_F(\gamma_L, \gamma_F)$ denote the expected values of $Q_L(x; \gamma_L(\eta_L), \gamma_F(\eta_F))$ and $Q_F(x; \gamma_L(\eta_L), \gamma_F(\eta_F))$, respectively, with expectations taken over the statistics of (x, y, z). This thus defines the normal form of the underlying game over the product strategy space $\Gamma_L \times \Gamma_F$. The Stackelberg solution one seeks is a pair $(\gamma_L^* \in \Gamma_L \gamma_F^* \in \Gamma_F)$ that satisfies:

$$\sup_{\gamma_F \in R_F(\gamma^*)} J_L(\gamma_L^*, \gamma_F) = \inf_{\gamma_L \in \Gamma_L} \sup_{\gamma_F \in R_F(\gamma_L)} J_L(\gamma_L, \gamma_F) \,.$$
(II.24)

where $R_F(\gamma_L)$ is the optimum reaction set of the follower to L's policy γ_L , that is

$$R_F(\gamma_L) = \{\gamma_F \in \Gamma_F : \gamma_F = \arg\min_{\beta \in \Gamma_F} J_F(\gamma_L, \beta)\},$$
(II.25)

and $\gamma_F^* \in R_F(\gamma_L^*)$. This is a *pessimistic* version of the Stackelberg equilibrium, where L takes the worst element out of the reaction set R_F , which is the most appropriate one if this set is not a singleton. The other extreme would be the *optimistic* one where in (II.24) sup is replaced by inf, which relies too much on a cooperative behavior by F. For a large class of problems, however, as also partially discussed below, and fully in [6], $R_F(\gamma_L^*)$ is a singleton, and hence these different views on the Stackelberg solution do not arise. Particularly, let us assume that all variables take values in appropriate-dimensional Euclidean spaces, and the

cost functions Q_L and Q_F are each jointly continuous and strictly convex in the pair (u_l, u_F) (as a special case, and more in line with the general theme of this paper, we can take Q_L and Q_F to be general quadratic in their arguments).

Direct Approach:

Assuming that $R_F(\gamma_L)$ is a singleton for each $\gamma_L \in \Gamma_L$, a direct approach toward obtaining the Stackelberg solution would entail the following steps:

- (i) Obtain the unique optimum response by F to γ_L : $R_F(\gamma_L)$.
- (ii) Minimize $J_L(\gamma_L, R_F(\gamma_L))$ over $\gamma_l \in \Gamma_L$: unique minimizer γ_L^* .
- (iii) Combining (i) and (ii), the Stackelberg equilibrium is $(\gamma_L^*, \gamma_F^* = R_F(\gamma_L^*))$.

The *challenge* in this direct approach lies in the computation of $R_F(\gamma_L)$, since $J_F(\gamma_L, \gamma_F)$ is structurally dependent on γ_L (for example, even with quadratic Q_F if the dependence of $\gamma_L(\eta_L)$ on u_F is nonlinear, minimization in (i) will no longer be quadratic; it could further be discontinuous in u_F , which would lead to further difficulties in the minimization of J_F). Hence, this direct approach, though looking simple and straightforward on the surface, meets unsurmountable difficulties, necessitating the development of an indirect approach, as discussed next.

Indirect Approach:

We first note an obvious but important and consequential observation regarding a lower bound on the Stackelberg cost of L. Toward this end, let us first introduce the static information set $\tilde{\eta}_L = \{y, z\}$, and denote the corresponding generic policy for L using this static information by f_L . and the corresponding policy space by F_L . Then, we have

$$\inf_{\gamma_L \in \Gamma_L, \gamma_F \in \Gamma_F} J_L(\gamma_L, \gamma_F) = \inf_{f_L \in F_L, \gamma_F \in \Gamma_F} J_L(f_L, \gamma_F) \\
\leq J_L(\gamma_L^*, \gamma_F^*), \quad (II.26)$$

where the equality follows because the minimum value of J_L with full cooperation of the two agents is the same regardless of whether L has access to u_F or not, since L has access to the measurement of F. In the case of a quadratic, strictly convex Q_L and with zero-mean Gaussian statistics, this cooperative (team) optimization problem admits a unique minimizing solution, where minimizing f_L is linear in (y.z) and γ_F linear in z. We now have the following steps that lead to the Stackelberg equilibrium in view of the property (II.26):

- (i') Find the optimum (team) performance for L with full cooperation of F over the space of policies $F_L \times \Gamma_F$: $(f_L^t, \gamma_F^t) = \arg \min_{(f_L, \gamma_F)} J_L(f_L, \gamma_F).$
- (ii') This generates a rich set of policies (γ_L^t, γ_F^t) in $\Gamma_L \times \Gamma_F$ leading to the same cost value for L, where γ_L^t satisfies the side condition $\gamma_L^t(y, z; u_F = \gamma_F^t) = f_L^t(y, z)$, with one such class of policies being $\gamma_L(y, z; u_F) = f_L^t(y, z) + g(u_f f_F^t(z))$, where g vanishes at the origin.
- (iii') Finding such a g then completes the solution to the Stackelberg game, where using the incentive design $\gamma_L(y, z; u_F)$ induces F to respond in such a way that

his optimum policy, γ_F^t , turns out to be the one most favorable to L.

It has been shown in [6] that such an incentive design with g taken as a linear function (or linear operator) exists for a large class of problems (even those where Q_L and Q_F are not necessarily quadratic), where the main tool used is the Hahn-Banach Supporting Hyperplane Theorem. This approach and construction can also be extended to stochastic games with multiple hierarchies and partial dynamic information [11].

III. Advances in Linear-Quadratic Stochastic Differential Games (by Jun Moon)

Since the seminal paper of Fleming and Souganidis in [39], stochastic differential games have been playing a central role in mathematical control theory, as they can be applied to model the general decision-making process between interacting players under stochastic uncertainties. Two different types of stochastic differential games can be formulated depending on the roles of the interacting players. Specifically, when the interaction of the players can be described in a symmetric way, it is called the Nash differential game can be used to formulate the nonsymmetric leader-follower hierarchical decision-making process between the players.

This section provides an overview of some recent results on stochastic LQ Nash and Stackelberg differential games.

A. Stochastic LQ Nash Differential Games with Random Coefficients

Consider the linear stochastic differential equation driven by the Wiener process $W = (W_1, \ldots, W_p)$:

$$dx(t) = \left[A(t)x(t) + B_1(t)u_1(t) + B_2(t)u_2(t)\right]dt \quad \text{(III.1)} + \sum_{i=1}^{p} \left[\sigma_i(t) + D_i(t)x(t)\right]dW_i(t), \ t \in [0, T],$$

where $x \in \mathbb{R}^n$ is the state with $x(0) = a, u_1 \in \mathbb{R}^{m_1}$ is the control of Player 1 and $u_2 \in \mathbb{R}^{m_2}$ is the control of Player 2. In (III.1), the coefficients satisfy $A : [0,T] \times \Omega \to \mathbb{R}^{n \times n}$, $B_1 : [0,T] \times \Omega \to \mathbb{R}^{n \times m_1}$, $B_2 : [0,T] \times \Omega \to \mathbb{R}^{n \times m_2}$, and $D_i : [0,T] \times \Omega \to \mathbb{R}^{n \times n}$ for $i = 1, \ldots, p$, which are uniformly bounded in $(\omega, t) \in \Omega \times [0,T]$ and $\{\mathcal{F}_t\}_{t \ge 0}$ -adapted stochastic processes. Also, $\sigma_i \in \mathcal{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^n)$, $i = 1, \ldots, p$. The set of admissible controls for Player *i*, i = 1, 2, is defined by

$$\mathcal{U}_i := \{ u_i(\cdot) = F_i(\cdot)x(\cdot) + \psi_i(\cdot) \mid \\ F_i \in \mathcal{C}_{\mathcal{F}}([0,T]; \mathbb{R}^{m_i \times n}) \text{ and } \psi_i \in \mathcal{L}^2_{\mathcal{F}}([0,T]; \mathbb{R}^n) \}.$$

The diffusion term in (III.1) consists of both stateindependent and state-dependent parts. The state-independent part corresponds to $\sigma_i(t) dW_i(t)$, which can be viewed as additive noise in (III.1). The state-dependent part is $D_i(t)x(t)dW_i(t)$, which is state multiplicative noise. Note that from [122, Theorem 6.16, Chapter 1], for any $(u_1, u_2) \in$ $\mathcal{U}_1 \times \mathcal{U}_2$, (III.1) admits a unique strong solution for any fixed initial conditions. We consider here the zero-sum game setting with a quadratic objective functional to be minimized by Player 1 and maximized by Player 2:

$$J(u_1, u_2) = \frac{1}{2} \mathbb{E} \Big[|x(T)|_M^2$$
(III.2)
+ $\int_0^T \Big[|x(t)|_{S(t)}^2 + |u_1(t)|_{R_1(t)}^2 - |u_2(t)|_{R_2(t)}^2 \Big] dt \Big],$

where $M: \Omega \to \mathbb{S}_{\geq 0}^n$ is an uniformly bounded (in $\omega \in \Omega$) \mathcal{F}_T -measurable symmetric random matrix, and $S: [0,T] \times \Omega \to \mathbb{S}_{\geq 0}^n$, $R_1: [0,T] \times \Omega \to \mathbb{S}^{m_1}$, $R_2: [0,T] \times \Omega \to \mathbb{S}^{m_2}$ are symmetric uniformly bounded in $(\omega,t) \in \Omega \times [0,T]$ and $\{\mathcal{F}_t\}_{t\geq 0}$ -adapted matrix-valued stochastic processes. Assume $R_i \in \mathbb{S}_{>0}^{m_i}$, i = 1, 2, i.e., they are symmetric uniformly positive definite for $(\omega, t) \in \Omega \times [0,T]$.

As already mentioned, in (III.2), Player 1 chooses $u_1 \in U_1$ to minimize it, while Player 2 selects $u_2 \in U_2$ to maximize the same objective functional, making (III.2) the cost for Player 1 and the payoff for Player 2. We could refer to this problem as an LQ stochastic zero-sum differential game (LQ-SZSDG) with *random coefficients*, where the corresponding stochastic state equation includes both additive and state multiplicative noises.

The objective of this subsection is to characterize a *feed-back Nash equilibrium* (equivalently, *saddle-point equilib-rium*) of the LQ-SZSDG. Specifically, from [16], the pair $(\bar{u}_1, \bar{u}_2) \in \mathcal{U}_1 \times \mathcal{U}_2$ constitutes the Nash equilibrium of the LQ-SZSDG if it satisfies the following inequalities:

$$J(\bar{u}_1, u_2) \le J(\bar{u}_1, \bar{u}_2) \le J(u_1, \bar{u}_2), \quad (\text{III.3})$$

for any $u_1 \in U_1$ and $u_2 \in U_2$. Note that in (III.3), $J(\bar{u}_1, \bar{u}_2)$ is the optimal game value of the LQ-SZSDG (if it exists). We characterize the explicit Nash equilibrium and the optimal game value of the LQ-SZSDG.

We introduce the stochastic Riccati differential equation (SRDE): for $t \in [0, T]$,

$$\begin{cases} dP(t) = -[A^{\top}(t)P(t) + P(t)A(t) + S(t) \\ -P(t)B_{1}(t)R_{1}^{-1}(t)B_{1}^{\top}(t)P(t) \\ +P(t)B_{2}(t)R_{2}^{-1}(t)B_{2}^{\top}(t)P(t)]dt \\ -\sum_{i=1}^{p} [D_{i}^{\top}(t)P(t)D_{i}(t) + Q_{i}(t)D_{i}(t) \\ +D_{i}^{\top}(t)Q_{i}(t)]dt + \sum_{i=1}^{p} Q_{i}(t)dW_{i}(t), \\ P(T) = M. \end{cases}$$
(III.4)

The *n*-dimensional linear backward stochastic differential equation (BSDE) with random coefficients is (for $t \in [0, T]$)

$$\begin{cases} ds(t) = -[A(t) - B_1(t)R_1^{-1}(t)B_1^{\top}(t)P(t) \\ +B_2(t)R_2^{-1}(t)B_2^{\top}(t)P(t)]^{\top}s(t)dt \\ -\sum_{i=1}^p [D_i^{\top}(t)P(t)\sigma_i(t) + Q_i(t)\sigma_i(t) \\ +D_i^{\top}(t)r_i(t)]dt + \sum_{i=1}^p r_i(t)dW_i(t), \\ s(T) = 0. \end{cases}$$
(III.5)

Theorem III.1 Consider the LQ-SZSDG with random coefficients with (III.2) and (III.1). Suppose that $(P,Q_1,\ldots,Q_p) \in C_{\mathcal{F}}([0,T];\mathbb{S}^n_{\geq 0}) \times \mathcal{L}^2_{\mathcal{F}}([0,T];\mathbb{S}^n,\ldots,\mathbb{S}^n)$ is the solution of the SRDE in (III.4) and $(s,r_1,\ldots,r_p) \in$

 $\mathcal{L}^{2}_{\mathcal{F}}(\Omega; C([0,T]; \mathbb{R}^{n})) \times \mathcal{L}^{2}_{\mathcal{F}}([0,T]; \mathbb{R}^{n}, \dots, \mathbb{R}^{n})$ is the solution of the BSDE in (III.5). Then the Nash equilibrium, $(u_{1}^{*}, u_{2}^{*}) \in \mathcal{U}_{1} \times \mathcal{U}_{2}$, satisfying (III.3), can be written as

$$\begin{cases} u_1^*(t) = -R_1^{-1}(t)B_1^{\top}(t)P(t)x(t) - R_1^{-1}(t)B_1^{\top}(t)s(t), \\ u_2^*(t) = R_2^{-1}(t)B_2^{\top}(t)P(t)x(t) + R_2^{-1}(t)B_2^{\top}(t)s(t). \end{cases}$$
(III.6)

Moreover, the corresponding optimal game value of the LQ-SZSDG under (III.6) is given by

$$J(u_1^*, u_2^*) = \mathbb{E}\Big[\frac{1}{2}|a|_{P(0)}^2 + a^{\top}s(0) + \int_0^T \Lambda(t)\mathrm{d}t\Big], \quad \text{(III.7)}$$

where
$$\begin{split} & \Lambda(t) &:= \frac{1}{2} \sum_{i=1}^{p} \sigma_{i}^{\top}(t) P(t) \sigma_{i}(t) &+ \\ & \sum_{i=1}^{p} \sigma_{i}^{\top}(t) r_{i}(t) &- \frac{1}{2} s^{\top}(t) B_{1}(t) R_{1}^{-1}(t) B_{1}^{\top}(t) s(t) &+ \\ & \frac{1}{2} s^{\top}(t) B_{2}(t) R_{2}^{-1}(t) B_{2}^{\top}(t) s(t). \end{split}$$

Note that (III.5) is a linear BSDE, which admits a unique solution due to [122, Theorem 2.2, Chapter 7], provided that the SRDE in (III.4) admits a unique solution. For the solvability of the SRDE in (III.4), we have the following result from [52, Proposition 2.1]:

Proposition III.1 Suppose that $(B_1(t)R_1^{-1}(t)B_1^{\top}(t) - B_2(t)R_2^{-1}(t)B_2^{\top}(t))$ is uniformly positive definite for $(\omega, t) \in \Omega \times [0, T]$, i.e., $(B_1(t)R_1^{-1}(t)B_1^{\top}(t) - B_2(t)R_2^{-1}(t)B_2^{\top}(t)) > 0$ for $(\omega, t) \in \Omega \times [0, T]$. Then (III.4) admits a unique solution with $(P, Q_1, \ldots, Q_p) \in \mathcal{C}_{\mathcal{F}}([0, T]; \mathbb{S}_{>0}^n) \times \mathcal{L}_{\mathcal{F}}^2([0, T]; \mathbb{S}^n, \ldots, \mathbb{S}^n).$

- **Remark III.1** (i) If all the coefficients in (III.1) and (III.2) are deterministic, then it is easy to see that $Q_i = r_i = 0$. In this case, the SRDE corresponds to the deterministic RDE in LQ stochastic differential games with deterministic coefficients studied in [68], [94].
- (ii) Let $R_2(t) = \mu^2 I$ with $\mu > 0$. Then as $\mu \to \infty$, Theorem III.1 is reduced to the one-player stochastic optimal control problem with random coefficients in [101, Theorem 3.2] and [27]. Under this limit, the SRDE admits a unique solution in view of Proposition III.1.
- (iii) The condition in Proposition III.1 can be checked easily, since it depends only on the system and cost parameters. To solve (III.4) and (III.5), numerical computation is essential. There are various numerical approaches for solving BSDEs.

More detailed results on the LQ differential game studied in this subsection can be found in [69].

B. Stochastic LQ Nash Differential Games for Mean-Field Type Systems

Stochastic differential equations (SDEs), in which the expected values of state and/or control variables are included in the drift and diffusion terms, constitute a class of mean field SDEs (MFSDEs). The theory of MFSDEs can be traced back to the study of McKean-Vlasov SDEs, a kind of MFSDEs, for analyzing interacting large-scale particle systems at the macroscopic level [49], [102]. Since then, there have been

concerted efforts on studying McKean-Vlasov SDEs and their applications [31], [90].

Stochastic optimal control for MFSDEs and their applications have been studied extensively in the literature. As mentioned in [26], [36], [119], the purpose of studying stochastic optimal control for MFSDEs is to analyze macroscopic behavior of large-scale interacting multi-agent systems and reduce the impact of random effects on the controlled state process. A complete solution to linear-quadratic (LQ) stochastic optimal control for MFSDEs was obtained in [119], where the linear feedback-type optimal solution was obtained in terms of a Riccati differential equation. The time-consistent optimal solution for LQ stochastic optimal control of MFSDEs was characterized in [121]. Recently, the linear-exponential-quadratic control problem for MFSDEs was considered in [82], and the time-inconsistent mean-field Stackelberg differential game was studied in [83].

Below, we provide a summary of the results in [70], which considered the LQ mean-field (MF) zero-sum differential game (ZSDG).

For a precise formulation of the ZSDG, consider the following linear SDE driven by a one-dimensional Brownian motion W:

$$dx(s) = \left[A_1(s)x(s) + A_2(s)\mathbb{E}[x(s)] + B_{11}(s)u_1(s) + B_{12}(s)\mathbb{E}[u_1(s)] + B_{21}(s)u_2(s) + B_{22}(s)\mathbb{E}[u_2(s)] \right] ds + \left[C_1(s)x(s) + C_2(s)\mathbb{E}[x(s)] + D_{12}(s)\mathbb{E}[u_1(s)] + D_{11}(s)u_1(s) + D_{22}(s)u_2(s) + D_{21}(s)\mathbb{E}[u_2(s)] \right] dW(s), \ s \in [0, T], .$$
(III.8)

Here $x \in \mathbb{R}^n$ is state with $x(0) = a, u_1 \in \mathbb{R}^{m_1}$ is control of Player 1 and $u_2 \in \mathbb{R}^{m_2}$ is control of Player 2; Wis the one-dimensional standard Brownian motion defined on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, where $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration generated by the Brownian motion; and $A_i(\cdot), B_{ij}(\cdot), C_i(\cdot)$ and $D_{ij}(\cdot)$ with i, j = 1, 2 are the coefficient matrices with appropriate dimensions, which are deterministic and continuous functions on [0, T].

In (III.8), the expected values of the state and the controls are included, which are known as *mean field* of the state and the controls. Hence, (III.8) can be regarded as a class of mean field stochastic differential equations (MFSDEs), which has been studied extensively in the literature, particularly, for reducing variation of random effects on the controlled process and macroscopic analysis of large-scale multi-agent systems.

The set of admissible controls for Player i, i = 1, 2, is defined by

$$\begin{aligned} \mathcal{U}_i &= \{ u_i : [0,T] \times \Omega \to \mathbb{R}^{m_i} : u_i(\cdot) \text{ is an} \\ &\{ \mathcal{F}_t \}_{t \ge 0} \text{-adapted process with } \mathbb{E} \int_0^T |u_i(s)|^2 \mathrm{d}s \}. \end{aligned}$$

Note that in view of [119, Proposition 2.6], for any $u_1 \in \mathcal{U}_1$

and $u_2 \in \mathcal{U}_2$, the MFSDE (III.8) admits a unique solution satisfying $\mathbb{E}[\sup_{s \in [0,T]} |x(s)|^2] < \infty$.

The quadratic objective functional for Players 1 and 2, as minimizer and maximizer, respectively, is given by

$$\begin{aligned} J(u_1, u_2) & \text{(III.9)} \\ &= \frac{1}{2} \mathbb{E} \Big[|x(T)|_{M_1}^2 + |\mathbb{E}[x(T)]|_{M_2}^2 + \int_0^T \big[|x(s)|_{Q_1(s)}^2 \\ &+ |\mathbb{E}[x(s)]|_{Q_2(s)}^2 + |u_1(s)|_{R_{11}(s)}^2 + |\mathbb{E}[u_1(s)]|_{R_{12}(s)}^2 \\ &+ |u_2(s)|_{R_{21}(s)}^2 + |\mathbb{E}[u_2(s)]|_{R_{22}(s)}^2 \big] \mathrm{d}s \Big], \end{aligned}$$

where $Q_i(\cdot), M_i \in \mathbb{S}^n$ and $R_{ij}(\cdot) \in \mathbb{S}^{m_i}$ for i, j = 1, 2. In (III.9), $Q_i(\cdot), M_i$ and $R_{ij}(\cdot)$ are the weighting matrices with appropriate dimensions, which are deterministic continuous (and therefore bounded) functions on [0, T]. We note that the weighting matrices need not be (positive and negative) definite. For this zero-sum game, Player 1 minimizes (III.9) by choosing u_1 , while Player 2 maximizes the same by selecting u_2 . Then the problem corresponds to the linear-quadratic mean field stochastic zero-sum differential game (LQ-MF-SZSDG).

The main objective here is to obtain a (feedback) saddlepoint equilibrium (equivalent to the (feedback) Nash equilibrium) for the LQ-MF-SZSDG. That is, the saddle-point equilibrium $(\bar{u}_1, \bar{u}_2) \in \mathcal{U}_1 \times \mathcal{U}_2$ of the LQ-MF-SZSDG satisfies the following pair of inequalities [16]:

$$J(\bar{u}_1, u_2) \le J(\bar{u}_1, \bar{u}_2) \le J(u_1, \bar{u}_2), \quad (\text{III.10})$$

for any $u_1 \in \mathcal{U}_1$ and $u_2 \in \mathcal{U}_2$. In this case, $J(\bar{u}_1, \bar{u}_2)$ is the saddle-point value of the LQ-MF-SZSDG.

We first introduce the following notation:

$$\begin{split} B_i(\cdot) &= \begin{bmatrix} B_{1i}(\cdot) & B_{2i}(\cdot) \end{bmatrix}, \ i = 1,2 \\ D_1(\cdot) &= \begin{bmatrix} D_{11}(\cdot) & D_{22}(\cdot) \end{bmatrix} \\ D_2(\cdot) &= \begin{bmatrix} D_{12}(\cdot) & D_{21}(\cdot) \end{bmatrix} \\ R_i(\cdot) &= \text{diag}\{R_{1i}(\cdot), \ R_{2i}(\cdot)\}, \ i = 1,2 \\ \bar{A}(\cdot) &= A_1(\cdot) + A_2(\cdot), \ \bar{Q}(\cdot) &= Q_1(\cdot) + Q_2(\cdot) \\ \bar{B}_i(\cdot) &= B_{i1}(\cdot) + B_{i2}(\cdot), \ i = 1,2 \\ \bar{D}_1(\cdot) &= D_{11}(\cdot) + D_{12}(\cdot) \ \bar{D}_2(\cdot) &= D_{22}(\cdot) + D_{21}(\cdot) \\ \bar{B}(\cdot) &= \begin{bmatrix} \bar{B}_1(\cdot) & \bar{B}_2(\cdot) \end{bmatrix} \\ \bar{D}(\cdot) &= \begin{bmatrix} \bar{D}_1(\cdot) & \bar{D}_2(\cdot) \end{bmatrix} \\ \bar{D}(\cdot) &= \begin{bmatrix} \bar{D}_1(\cdot) & \bar{D}_2(\cdot) \end{bmatrix} \\ \bar{C}(\cdot) &= C_1(\cdot) + C_2(\cdot), \ \bar{M} = M_1 + M_2 \\ \bar{R}_i(\cdot) &= R_{i1}(\cdot) + R_{i2}(\cdot), \ i = 1,2 \\ \bar{R}(\cdot) &= \text{diag}\{\bar{R}_1(\cdot), \ \bar{R}_2(\cdot)\}. \end{split}$$

Consider the following coupled Riccati differential equations

$$(\text{CRDEs}): \text{ for } s \in [0, T], \\ \begin{cases} -\frac{\mathrm{d}P(s)}{\mathrm{d}s} = A_1^{\top}(s)P(s) + P(s)A_1(s) + Q_1(s) \\ +C_1^{\top}(s)P(s)C_1(s) \\ -\left[(P(s)B_1(s) + C_1^{\top}(s)P(s)D_1(s)) \\ \times (R_1(s) + D_1^{\top}(s)P(s)D_1(s))^{-1} \\ \times (P(s)B_1(s) + C_1^{\top}(s)P(s)D_1(s))^{\top} \right], \\ P(T) = M_1, \end{cases}$$
(III.11)

and for $s \in [0, T]$,

$$\begin{cases} -\frac{\mathrm{d}Z(s)}{\mathrm{d}s} = \bar{A}^{\top}(s)Z(s) + Z(s)\bar{A}(s) + \bar{Q}(s) \\ +\bar{C}^{\top}(s)P(s)\bar{C}(s) \\ -\left[(Z(s)\bar{B}(s) + \bar{C}(s)P(s)\bar{D}(s)) \\ \times(\bar{R}(s) + \bar{D}^{\top}(s)P(s)\bar{D}(s))^{-1} \\ \times(Z(s)\bar{B}(s) + \bar{C}(s)P(s)\bar{D}(s))^{\top} \right], \end{cases}$$
(III.12)
$$Z(T) = \bar{M}.$$

Note that this is a one-way coupling, with RDE for Z in (III.12) depending on P, but the RDE for P in (III.11) not depending on Z. It can be seen that $P, Z \in \mathbb{S}^n$. Let

$$\Lambda_{1}(\cdot) = (B_{11}^{\top}(\cdot)P(\cdot) + D_{11}^{\top}(\cdot)P(\cdot)C_{1}(\cdot))^{\top}$$

$$\Lambda_{2}(\cdot) = (B_{21}^{\top}(\cdot)P(\cdot) + D_{22}^{\top}(\cdot)P(\cdot)C_{1}(\cdot))^{\top}$$

$$\Phi(\cdot) = \begin{bmatrix} \Phi_{11}(\cdot) & \Phi_{12}(\cdot) \\ \Phi_{12}^{\top}(\cdot) & \Phi_{22}(\cdot) \end{bmatrix},$$

where $\Phi_{11}(\cdot) = R_{11}(\cdot) + D_{11}^{\top}(\cdot)P(\cdot)D_{11}(\cdot), \ \Phi_{22}(\cdot) = R_{21}(\cdot) + D_{22}^{\top}(\cdot)P(\cdot)D_{22}(\cdot) \text{ and } \Phi_{12}(\cdot) = D_{11}^{\top}(\cdot)P(\cdot)D_{22}(\cdot).$ Similarly, we define

$$\begin{split} \bar{\Lambda}_1(\cdot) &= (\bar{B}_1^\top(\cdot)Z(\cdot) + \bar{D}_1^\top(\cdot)P(\cdot)\bar{C}(\cdot))^\top \\ \bar{\Lambda}_2(\cdot) &= (\bar{B}_2^\top(\cdot)Z(\cdot) + \bar{D}_2^\top(\cdot)P(\cdot)\bar{C}(\cdot))^\top \\ \bar{\Phi}(\cdot) &= \begin{bmatrix} \bar{\Phi}_{11}(\cdot) & \bar{\Phi}_{12}(\cdot) \\ \bar{\Phi}_{12}^\top(\cdot) & \bar{\Phi}_{22}(\cdot) \end{bmatrix}, \end{split}$$

where $\bar{\Phi}_{11}(\cdot) = \bar{R}_1(\cdot) + \bar{D}_1^\top(\cdot)P(\cdot)\bar{D}_1(\cdot), \ \bar{\Phi}_{22}(\cdot) = \bar{R}_2(\cdot) + \bar{D}_2^\top(\cdot)P(\cdot)\bar{D}_2(\cdot) \ \text{and} \ \bar{\Phi}_{12}(\cdot) = \bar{D}_1^\top(\cdot)P(\cdot)\bar{D}_2(\cdot).$

Theorem III.2 Assume that the CRDEs (III.11) and (III.12) admit solutions on [0,T]. Suppose that $\Phi_{11}(t) > 0$, $\bar{\Phi}_{11}(t) > 0$, $\Phi_{22}(t) < 0$ and $\bar{\Phi}_{22}(t) < 0$ for $t \in [0,T]$. Then the pair (u_1^*, u_2^*) given by

$$\begin{cases} u_{1}^{*}(t) = -\left[\Phi_{11}(t) - \Phi_{12}(t)\Phi_{22}^{-1}(t)\Phi_{12}^{\top}(t)\right]^{-1}\left[\Lambda_{1}^{\top}(t) - \Phi_{12}(t)\Phi_{22}^{-1}(t)\Lambda_{2}^{\top}(t)\right](x(t) - \mathbb{E}[x(t)]) \\ -\left[\bar{\Phi}_{11}(t) - \bar{\Phi}_{12}(t)\bar{\Phi}_{22}^{-1}(t)\bar{\Phi}_{12}^{\top}(t)\right]^{-1}\left[\bar{\Lambda}_{1}^{\top}(t) - \bar{\Phi}_{12}(t)\bar{\Phi}_{22}^{-1}(t)\bar{\Phi}_{12}^{\top}(t)\right] \right] \\ -\bar{\Phi}_{12}(t)\bar{\Phi}_{22}^{-1}(t)\bar{\Lambda}_{2}^{\top}(t)\right]\mathbb{E}[x(t)], \\ u_{2}^{*}(t) = -\left[\Phi_{22}(t) - \Phi_{12}^{\top}(t)\Phi_{11}^{-1}(t)\Phi_{12}(t)\right]^{-1}\left[\Lambda_{2}^{\top}(t) - \Phi_{12}^{\top}(t)\Phi_{11}^{-1}(t)\bar{\Phi}_{12}(t)\right] \\ -\bar{\Phi}_{12}^{\top}(t)\Phi_{11}^{-1}(t)\bar{\Lambda}_{1}^{\top}(t)\right](x(t) - \mathbb{E}[x(t)]) \\ -\left[\bar{\Phi}_{22}(t) - \bar{\Phi}_{12}^{\top}(t)\bar{\Phi}_{11}^{-1}(t)\bar{\Phi}_{12}(t)\right]^{-1}\left[\bar{\Lambda}_{2}^{\top}(t) - \bar{\Phi}_{12}^{\top}(t)\bar{\Phi}_{11}^{-1}(t)\bar{\Lambda}_{1}^{\top}(t)\right]\mathbb{E}[x(t)], \end{cases}$$

constitutes a saddle-point (Nash) equilibrium of the LQ-MF-SZSDG, i.e., the pair (u_1^*, u_2^*) satisfies the inequalities in (III.10). Also, the corresponding saddle-point value is $J(u_1^*, u_2^*) = \frac{1}{2}a^\top Z(0)a.$

A detailed analysis of this problem can be found in [70]. In addition, several different formulations of stochastic LQ Nash differential games have been studied in [67], [68], [70], [82], [94], [96], [99], [124] and the references therein.

C. Stochastic LQ Stackelberg Differential Games for Jump-Diffusion Models

We now consider here a hierarchical formulation of an LQ stochastic differential game, with one leader and one follower. Let us have the following controlled stochastic differential equation on [t, T] driven by both a one-dimensional Brownian motion B and a (compensated) Poisson jump-diffusion process \tilde{N} :

$$\begin{cases} dx(s) = \left[A(s)x(s-) + B_1(s)u_1(s) + B_2(s)u_2(s)\right] ds \\ + \left[C(s)x(s-) + D_1(s)u_1(s) + D_2(s)u_2(s)\right] dB(s) \\ + \int_E \left[F(s,e)x(s-) + G_1(s,e)u_1(s) \\ + G_2(s,e)u_2(s)\right] \widetilde{N}(de,ds), \ s \in [t,T], \\ x(t) = a, \end{cases}$$
(III.13)

where, $x \in \mathbb{R}^n$ is the state process, $u_1 \in \mathbb{R}^{m_1}$ is the control of the leader, and $u_2 \in \mathbb{R}^{m_2}$ is the control of the follower. Let $\mathcal{U}_1 := \mathcal{L}^2_{\mathbb{F},p}(t,T;\mathbb{R}^{m_1})$ and $\mathcal{U}_2 :=$ $\mathcal{L}^2_{\mathbb{F},p}(t,T;\mathbb{R}^{m_2})$ be spaces of admissible controls for the leader and the follower, respectively. Note that A, C : $\Omega \times [0,T] \to \mathbb{R}^{n \times n}, B_i, D_i : \Omega \times [0,T] \to \mathbb{R}^{n \times m_i},$ $i = 1,2, F : \Omega \times [0,T] \to G^2(E,\mathcal{B}(E),\lambda;\mathbb{R}^{n \times n}),$ $G_i : \Omega \times [0,T] \to G^2(E,\mathcal{B}(E),\lambda;\mathbb{R}^{n \times m_i}), i = 1,2,$ are \mathcal{F}_s -predictable stochastic processes (random coefficients of (III.13)), which are continuous in $t \in [0,T]$ and uniformly bounded in a.e $(\omega, t) \in \Omega \times [0,T]$. We note that for any $(u_1, u_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, (III.13) admits a unique (strong) càdlàg solution in $\mathcal{C}^2_{\mathbb{F}}(t,T;\mathbb{R}^n)$.

The objective functional to be minimized by the leader is given by

$$J_1(a; u_1, u_2) = \mathbb{E}\left[\int_t^T \left[|x(s)|^2_{Q_1(s)} + |u_1(s)|^2_{R_1(s)}\right] \mathrm{d}s + |x(T)|^2_{M_1}\right],$$
 (III.14)

and the objective functional of the follower is:

$$J_{2}(a; u_{1}, u_{2}) = \mathbb{E}\left[\int_{t}^{T} \left[|x(s)|^{2}_{Q_{2}(s)} + |u_{2}(s)|^{2}_{R_{2}(s)}\right] \mathrm{d}s + |x(T)|^{2}_{M_{2}}\right].$$
 (III.15)

Here, $Q_i : \Omega \times [0,T] \to \mathbb{S}^n$ and $R_i : \Omega \times [0,T] \to \mathbb{S}^{m_i}$ i = 1, 2, are \mathcal{F}_s -predictable stochastic processes (random coefficients of (III.14) and (III.15)), which are continuous in $t \in [0,T]$ and uniformly bounded in a.e $(\omega, t) \in \Omega \times [0,T]$. Also, $M_i : \Omega \to \mathbb{S}^n$, i = 1, 2, are \mathcal{F}_T -measurable random matrices, which are uniformly bounded in a.e. $\omega \in \Omega$. The interaction between the leader and the follower in this LQ Stackelberg game can be described as follows. The leader chooses and announces her (or his) optimal solution to the follower by considering the rational reaction of the follower. The follower then determines his (or her) optimal solution by responding to the optimal solution of the leader. We refer to this problem as the *linear-quadratic (LQ) stochastic Stackelberg differential game for jump-diffusion systems with random coefficients*.

Under this setting, the problem can be solved in a reverse way [16], [25], [123]. Specifically, the main objective of the follower is to minimize (III.15) subject to (III.13) for any control of the leader $u_1 \in U_1$, i.e., for any $u_1 \in U_1$,

(LQ-F)
$$J_2(a; u_1, \overline{u}_2[a, u_1]) = \inf_{u_2 \in \mathcal{U}_2} J_2(a; u_1, u_2).$$

We note that from (LQ-F), \overline{u}_2 is an optimal strategy dependent on $(a, u_1) \in \mathbb{R}^n \times \mathcal{U}_1$. Then given the optimal solution of (LQ-F), the problem of the leader can be stated as follows:

(LQ-L)
$$J_1(a; \overline{u}_1, \overline{u}_2[a, \overline{u}_1]) = \inf_{u_1 \in \mathcal{U}_1} J_1(a; u_1, \overline{u}_2[a, u_1])$$

When the pair $(\overline{u}_1, \overline{u}_2[a, \overline{u}_1]) \in \mathcal{U}_1 \times \mathcal{U}_2$ in **(LQ-F)** and **(LQ-L)** exists, we say that the pair $(\overline{u}_1, \overline{u}_2[a, \overline{u}_1])$ constitutes an (adapted) open-loop type *Stackelberg equilibrium* for in the Stackelberg game [16], [25], [83], [123].

A complete analysis of (LQ-F) and (LQ-L) is given in [72]. Below, we discuss the main challenges and approaches.

As discussed above, we first need to solve (LQ-F). In particular, using the stochastic maximum principle for jumpdiffusion systems, we obtain an open-loop type optimal solution for (LQ-F) in terms of the forward-backward stochastic differential equation (FBSDE) with jump diffusions and random coefficients, which explicitly depends on $(a, u_1) \in$ $\mathbb{R}^n \times \mathcal{U}_1$. Since the open-loop type optimal solution is not implementable in practical situations, we have to obtain its state-feedback representation in terms of the integrostochastic Riccati differential equation (ISRDE) by extending the Four-Step Scheme of [123] to the case of general jumpdiffusion models. Then it is necessary to show that the corresponding state-feedback type control is the optimal solution for (LQ-F) via the completion of squares method.

Using the optimal solution of (LQ-F), the second step is to solve (LQ-L). (LQ-L) is the (indefinite) LQ stochastic optimal control problem for FBSDEs with jump diffusions and random coefficients, where the FBSDE constraint, induced from (LQ-F), characterizes the rational reaction behavior of the follower [16], [25]. It is necessary to obtain the stochastic maximum principle for (LQ-L) using the variational approach and duality analysis. Then by the stochastic maximum principle, the open-loop optimal solution for (LQ-L) should be obtained in terms of the coupled FBSDEs with jump diffusions and random coefficients.

The state-feedback representation of the open-loop optimal solution of (LQ-L) in terms of the ISRDE can be obtained by establishing the Four-Step Scheme for FBSDEs with jump diffusions and random coefficients. Unfortunately, as discussed in [72], there is a technical limitation, which did not appear in [123]. This technical challenge arises due to the coupling structure in the Four-Step Scheme between the Brownian motion and the jump-diffusion process. Hence, we have to consider two different cases of (LQ-L):

- (i) the Poisson process N has jumps of unit size;
- (ii) the jump part of (III.13) does not depend on the control of the follower ($G_2 = 0$).

In [72], the above two different cases of (**LQ-L**) are considered, and their complete treatments are provided.

We note that other variations of stochastic LQ Stackelberg differential games including the classical results were studied in [16], [18], [25], [30], [74], [77], [80], [83], [123] and appropriate references therein.

IV. PROGRESS ON NONSTANDARD LQ CONTROL AND APPLICATIONS IN NCSS (BY HUANSHUI ZHANG)

In this section, we study LQ problems under nonstandard settings with (i) irregular condition, (ii) delay systems, and (iii) asymmetric information structure.

A. Why Is It Difficult?

The challenges mentioned in Section I show that the research on nonstandard LQ problems is still confronted with fundamental obstacles. In order to solve these problems, it is crucial to uncover the root causes of the complexity of these challenges while figuring out how to resolve them. Research by the author has shown that the root causes to the challenges, which the nonstandard LQ control faces, lie in problems in the decoupling solution of the state forward equation and the adjoint-state backward equation originated from the maximum principle. The main basis is: the solution for optimal control includes two parts: 1) deriving the maximum principle, that is, the state equation is used as a dynamic constraint and the problem is translated into an unconstrained quadratic optimization problem by using the dynamic Lagrangian product factor, to get the equilibrium condition and the backward equation satisfied with the adjoint state; 2) decoupling forward and backward differential/difference equations (FBDEs), that is, the adjoint state is expressed in the form of a state, and then the solution in the feedback form is obtained by using the equilibrium condition. However, the main obstacle for nonstandard LQ (irregular, infinite-dimensional and asymmetric) control lies in the 2nd step mentioned above, decoupling FBDEs. The following explains it in detail in terms of irregularity, infinite dimension and asymmetry.

1) Irregular LQ control: Consider the continuous-time linear system:

$$\dot{x}(t) = Ax(t) + Bu(t), \ x(t_0) = x_0,$$
 (IV.1)

and the quadratic performance:

$$J_T(t_0, x_0; u) = \int_{t_0}^T [x'(t)Qx(t) + u'(t)Ru(t)]dt + x'(T)Hx(T),$$
(IV.2)

where $Q \ge 0, R \ge 0, H \ge 0$.

Minimizing $J_T(t_0, x_0; u)$, u(t) satisfies the following equilibrium condition (maximum principle):

$$0 = Ru(t) + B'p(t), \qquad (IV.3)$$

where the adjoint state p(t) satisfies the backward equation:

$$\dot{p}(t) = -[A'p(t) + Qx(t)], \ p(T) = Hx(T).$$
 (IV.4)

Define the following Riccati equation:

$$0 = \dot{P}(t) + A'P(t) + P(t)A + Q - P(t)BR^{\dagger}B'P(t)$$
 [IV.5)

with P(T) = H and R^{\dagger} being the pseudo-inverse of R. If

$$Range(B'P(t)) \subseteq Range(R),$$
 (IV.6)

the above-mentioned problem of optimal control is called regular and the solution of FBDEs (IV.1), (IV.3), (IV.4) is p(t) = P(t)x(t). Thus the controller is given by

$$u(t) = -R^{\dagger}B'P(t)x(t) + (I - R^{\dagger}R)z(t).$$

in which z(t) is any vector [50].

If

$$Range(B'P(t)) \not\subseteq Range(R),$$
 (IV.7)

then the corresponding optimal control is called irregular, and $p(t) \neq P(t)x(t)$. Obtaining the solutions to FBDEs (IV.1), (IV.3), (IV.4) faces challenges, which is the fundamental reason why the irregular LQ has not be solved for a long time although considerable efforts have been made since the 1970s.

In [42], [84], [110] and references therein, the singular LQ control was studied by using "Transformation in state space", where the problem with zero control weighting matrix (R =0) was studied. It was shown that the problem is solvable if the initial value is given like $x_2(0) = C_{21}(0)x_1(0)$. Otherwise, an impulse control must be applied at the initial time [43]. In other words, the approach of "Transformation in state space" is only applicable to the case of specified initial value. In [28], [53], [131], the approach of "higher order maximum principle" was applied to singular LQ control. However, if the higher derivatives vanish, it is impossible to find the singular control with this approach [41]. The third approach is the perturbation approach in [32], [92]. The optimal solution is obtained by using the limit of the solution to Riccati equation when the perturbation is approaching to zero.

2) Stochastic LQ control of time-delay system: Even though the control problem of time-delay systems has been well studied, especially for deterministic systems and stochastic systems with additive noise [89], [91], the stochastic LQ control with multiplicative noise remains unsolved. Consider a continuous-time multiplicative-noise stochastic system with time delay

$$dx(t) = [Ax(t) + Bu(t-h)]dt$$
(IV.8)
+ $[\bar{A}x(t) + \bar{B}u(t-h)]dw(t),$

and quadratic performance:

$$J = \mathbb{E}\left[\int_0^T x'(t)Qx(t)dt + \int_h^T u'(t-h)Ru(t-h)dt + x'(T)Hx(T)\right],$$

where $Q \ge 0, R \ge 0, H \ge 0$. In (IV.8), w is the standard Brownian motion.

Minimizing J, u(t) satisfies the following equilibrium condition (via the maximum principle):

$$0 = Ru(t) + \mathbb{E}[B'p(t+h) + \bar{B}'q(t+h)|\mathcal{F}_t]. \quad (IV.9)$$

where adjoint states p(t) and q(t) satisfy the following BSDE:

$$\begin{cases} \dot{p}(t) = -[A'p(t) + \bar{A'}q(t) + Qx(t)]dt + q(t)dw(t), \\ p(T) = Hx(T). \end{cases}$$
(IV.10)

If h = 0, the above-mentioned problem represents the standard stochastic LQ control, the solution of FBDEs (IV.8)-(IV.10) is p(t) = P(t)x(t). Thus the controller is given by $u(t) = -[R + B'P(t)B]^{-1}[B'P(t) + B'P(t)A]x(t)$ [33]. If $h \neq 0, p(t) \neq P(t)x(t)$. How to get solutions of FBDEs (IV.8)-(IV.10) faces challenges, which is the fundamental reason why the time-delay stochastic LQ control has been open for a long time.

3) Information-asymmetric LQ control: Consider the following discrete-time system with two control inputs

$$x_{k+1} = Ax_k + B_1 u_k^1 + B_2 u_k^2 + w_k, \qquad \text{(IV.11)}$$

two observations

$$y_k^1 = H_1 x_k + v_k^1, y_k^2 = H_2 x_k + v_k^2,$$
 (IV.12)

and quadratic performance

$$J_{N} = \mathbb{E} \left\{ \sum_{k=0}^{N} [x_{k}^{\prime} Q x_{k} + (u_{k}^{1})^{\prime} R_{1} u_{k}^{1} + (u_{k}^{2})^{\prime} R_{2} u_{k}^{2}] + x_{N+1}^{\prime} H x_{N+1} \right\}.$$
 (IV.13)

Based on above-described observation, define an information set: $\mathcal{F}\{Y_k^2\} = \mathcal{F}\{y_0^2, \dots, y_k^2, u_0^2, \dots, u_{k-1}^2\}, \ \mathcal{F}\{Y_k\} = \mathcal{F}\{y_0^i, \dots, y_k^i, u_0^i, \dots, u_{k-1}^i, i = 1, 2\},$ obviously $\mathcal{F}\{Y_k^2\} \subseteq \mathcal{F}\{Y_k\}$, called information inclusion pattern.

Assuming u_k^1 is $\mathcal{F}{Y_k}$ -adapted and u_k^2 is $\mathcal{F}{Y_k^2}$ adapted, and minimizing J_N , the control u satisfies the following stationarity condition

$$0 = \mathbb{E}[B_1'\lambda_k|\mathcal{F}\{Y_k\}] + R_1 u_k^1, \qquad (IV.14)$$

$$0 = \mathbb{E}[B'_{2}\lambda_{k}|\mathcal{F}\{Y^{2}_{k}\}] + R_{2}u^{2}_{k}, \quad (IV.15)$$

where the adjoint state λ_k satisfies the following backward equation

$$\lambda_{k-1} = \mathbb{E}[A'\lambda_k + Qx_k|\mathcal{F}\{Y_k\}],$$

$$\lambda_N = \mathbb{E}[Hx_{N+1}|\mathcal{F}\{Y_{N+1}\}], \qquad (IV.16)$$

The adaptability assumed for \boldsymbol{u}_k^1 and \boldsymbol{u}_k^2 is so different

that the above FBDEs are very complex. As a result, the decentralized control in the case of information inclusion has not been fundamentally solved although the special case with sharing control information has been solved in discrete time [44]. For some continuous-time results, see [17].

4) The key technique for nonstandard LQ control: For FBDEs aiming at the above nonstandard LQ control problems, we have proposed a general method for decoupling the solution process. The general idea is to: 1) use the backward iterative induction method to obtain the solutions of FBDEs in the discrete-time case; 2) use the method of discretizing the continuous-time system and the one of approximating results of discrete-time system to continuous-time case to resolve the problems in decoupling solution of continuous-time FBDEs. For detailed solution methods and results, please refer to the references [118], [114], [66].

Since the decoupling problem of FBDEs is resolved, problems in nonstandard LQ control of the type formulated here have been fundamentally solved. Details can be found further below as well as in [62], [128], [129].

B. Solutions to Nonstandard LQ Control

1) Irregular LQ control: The solution to irregular LQ control is given below.

Theorem IV.1 Irregular LQ Control Problem is solvable if and only if there exists a matrix $P_1(T)$ satisfying $0 = B'_0(T)[P(T) + P_1(T)]$ such that the following modified cost

$$\bar{J}_0(t_0, x_0; u) = J_0(t_0, x_0; u) + x'(T)P_1(T)x(T)$$
 (IV.17)

is regular and $P_1(T)x(T) = 0$ is achieved with the controller minimizing (IV.17).

If the problem is solvable, the control is given as

$$u(t) = -R^{\dagger}(t)\bar{P}(t)x(t) + [I - R(t)R^{\dagger}(t)]z(t),$$

where the first part is to minimize the cost function (IV.17), and z(t) is to guarantee $P_1(T)x(T) = 0$. Details can be found in [129].

The general cases with additive noise and multiplicative noise can be found in [115] and [130].

- **Remark IV.1** It is obvious that $P_1(T) = 0$ for the regular (standard) LQ control, while $P_1(T) \neq 0$ for the irregular LQ control. So an essential difference of irregular LQ from regular one is that the irregular controller (if exists) needs to do two things at the same time: one is to minimize the cost (IV.17) and the other is to achieve the terminal constraint $P_1(T)x(T) = 0$.
 - In spite of this difference, the LQ control problem (irregular and regular) can be solved in a unified way as in Theorem IV.1.

2) Stochastic LQ control with time delay: To present the solution, we define a differential Riccati-ZXL equation as

$$-\dot{Z}(t) = A'Z(t) + Z(t)A$$
(IV.18)
+ $\dot{A}'X(t)\bar{A} - L(t) + Q,$
$$X(t) = Z(t) + \int_{t}^{t+h} e^{A'(s-t)}L(s)e^{A(s-t)}ds,$$
(IV.19)

where

$$L(s) = K'(s)\Omega(s)K(s)$$
(IV.20)

$$\Omega(s) = R + \bar{B}'X(s)\bar{B},$$

$$K(s) = -\Omega^{-1}(s)[B'Z(s) + \bar{B}'X(s)\bar{A}],$$

with the terminal values Z(T) = P(T) and X(T) = P(T).

Theorem IV.2 Stochastic LQ control with time delay is uniquely solvable if and only if the differential Riccati-ZXL equation (IV.18)-(IV.19) admits a solution satisfying $\Omega(t) > 0$ for $h \le t \le T$. In this case, the optimal control is

$$u(t-h) = K(t)\hat{x}(t|t-h),$$
 (IV.21)

where

$$\hat{x}(t|t-h) = e^{Ah}x(t-h)$$
(IV.22)
+
$$\int_{t-h}^{t} e^{A(t-\theta)}Bu(\theta-h)d\theta,$$

and the matrices $\Omega(t)$ and K(t) are given by (IV-B.2). Moreover, the optimal cost is

$$J_T^* = \frac{1}{2} \mathbb{E} \Biggl\{ \int_0^h x'(t) Q x(t) dt$$
(IV.23)
+ $x'(h) \Bigl[P(h) x(h)$
- $\int_0^h e^{A'\theta} \Pi(h+\theta,h+\theta) e^{A\theta} \hat{x}(h|\theta) d\theta \Bigr] \Biggr\},$

where P(h) = X(h),

$$\hat{x}(h|\theta) = e^{A(h-\theta)}x(\theta) + \int_{\theta}^{h} e^{A(h-\tau)}Bu(\tau-h)d\tau$$
$$\Pi(h+\theta,h+\theta) = L(h+\theta).$$

Details of the above can be found in [128].

The general cases of stochastic LQ control with multiple input delays, state delay can be found in [58], [106], [66].

Remark IV.2 It is obvious that the result presented in Theorem IV.2 includes the traditional stochastic control (i.e., h = 0) and deterministic control (i.e., $\overline{A} = 0, \overline{B} = 0$) as special cases. It is noted that the traditional stochastic control problem has been extensively studied from 1960s due to its wide applications. Plenty of progress has been made, mainly including the stochastic maximum principle [107], the design of the LQ controller based on generalized Riccati equation [27] and the indefinite stochastic LQ control theory [33]. The stochastic LQ control problem with time delay has remained challenging due to the fact that the separation principle does not hold any more, but Theorem IV.2 provides the complete solution to it.

3) Information-asymmetric LQ control: To present the explicit solution, we define the following coupled Riccati equations:

$$P_{k} = A' P_{k+1} A - M'_{k} \Upsilon_{k}^{-1} M_{k} + Q, \qquad (IV.24)$$

$$S_k = A' \Phi_{k+1} A - L_k^{0'} \Lambda_k^{-1} L_k + Q, \qquad (IV.25)$$

where

$$M_k = B' P_{k+1} A, \tag{IV.26}$$

$$\Upsilon_k = B' P_{k+1} B + R, \qquad (IV.27)$$

$$\Phi_k = (P_k - S_k)G_{k|k-1}^2 H_2 + S_k, \qquad (IV.28)$$

$$L_k^0 = B_1' \Phi_{k+1}' A, (IV.29)$$

$$L_k = B_1' \Phi_{k+1} A, \tag{IV.30}$$

$$\Lambda_k = B_1' \Phi_{k+1} B_1 + R_1, \tag{IV.31}$$

with terminal values $P_{N+1} = S_{N+1} = \Theta$.

Theorem IV.3 Assuming that Υ_k and Λ_k are invertible for k = N, ..., 0, the optimal controllers for informationasymmetric LQ control are given by

$$u_k = -\Upsilon_k^{-1} M_k \hat{x}_{k|k}^2, \qquad (\text{IV.32})$$

$$\tilde{u}_{k}^{1} = -\Lambda_{k}^{-1} L_{k} (\hat{x}_{k|k}^{1} - \hat{x}_{k|k}^{2}), \qquad \text{(IV.33)}$$

where $\hat{x}_{k|k}^2$ and $\hat{x}_{k|k}^1$ are defined as

$$\hat{x}_{k|k}^{1} = \hat{x}_{k|k-1}^{1} + G_{k|k-1}(y_{k} - H\hat{x}_{k|k-1}^{1}), \quad (IV.34)$$

$$\hat{x}_{k|k-1}^{i} = A\hat{x}_{k-1|k-1}^{i} + Bu_{k-1} + B_1\tilde{u}_{k-1}^{i}, \qquad (IV.35)$$

$$\hat{x}_{k|k}^2 = \hat{x}_{k|k-1}^2 + G_{k|k-1}^2 (y_k^2 - H_2 \hat{x}_{k|k-1}^2), \quad \text{(IV.36)}$$

$$\hat{x}_{k|k-1}^2 = A\hat{x}_{k-1|k-1}^2 + Bu_{k-1}, \qquad (IV.37)$$

where $G_{k|k-1} = \sum_{k|k-1}^{1} H'(H\sum_{k|k-1}^{1} H' + Q_v)^{-1}, G_{k|k-1}^2 = \sum_{k|k-1}^{2} H'_2(H_2\sum_{k|k-1}^{2} H'_2 + Q_v)^{-1}$ and $\sum_{k|k-1}^{1}, \sum_{k|k-1}^{2}$ are the estimation error covariances, initial values are given by $x_{0|-1}^2 = \mu$, $\sum_{0|-1}^{2} = \sigma$, and $\Upsilon_k, M_k, \Lambda_k, L_k$ are as in (IV.24)-(IV.31). Accordingly, the optimal $u_k^1 = \begin{bmatrix} I & 0 \end{bmatrix} u_k + \tilde{u}_k^1$, and the optimal $u_k^2 = \begin{bmatrix} 0 & I \end{bmatrix} u_k$. The optimal cost is obtained as

$$J_{N}^{*} = \mathbb{E}[x_{0}'P_{0}\hat{x}_{0|0}^{2} + x_{0}'S_{0}(\hat{x}_{0|0}^{1} - \hat{x}_{0|0}^{2})] \quad (IV.38)$$

$$+ tr(\Sigma_{N+1|N+1}^{1}\Theta)$$

$$+ \sum_{k=0}^{N} tr\left(\Sigma_{k|k}^{1}[Q - A'(S_{k+1} - \Phi_{k+1} - S_{k+1})] + [Q_{\omega}(S_{k+1}G_{k+1|k}H + \Phi_{k+1} - S_{k+1})]\right).$$

Details of the above can be found in [62].

Remark IV.3 It is noted that even though the studied information structure satisfies $\mathcal{F}{Y_k^2} \subseteq \mathcal{F}{Y_k}$, the problem is completely different from that in the literature [44] because the controller information of u^1 is unknown for controller u^2 which leads to solvability challenge. Also, it is different from the results in [56] where the system is required to be with nested structure and the common-information approach [86]. Moreover, the LQG problem with d-step delayed information sharing pattern has been further solved in [117].

V. STOCHASTIC LINEAR-QUADRATIC OPTIMAL Control Problems – Some Recent Results (by Jiongmin Yong)

In this section, we provide a brief overview of some additional recent results on LQ control problems.

Consider the following controlled linear ordinary differential equation (ODE, for short):

$$\begin{cases} \dot{X}(t) = AX(t) + Bu(t), & t \in [0, T], \\ X(0) = x, \end{cases}$$
(V.1)

where $X(\cdot)$ is the *state trajectory* valued in \mathbb{R}^n with x being the *initial state*, and $u(\cdot)$ is a *control function* valued in \mathbb{R}^m ; $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ are given matrices. We refer to the above as the *state equation*. Clearly, for any $x \in \mathbb{R}^n$ and $u(\cdot) \in \mathcal{U}[0, T]$, the set of all square integrable functions (valued in \mathbb{R}^m), (V.1) admits a unique solution $X(\cdot) \equiv X(\cdot; x, u(\cdot))$.

To measure the performance of the control, one can introduce the following *cost functional*:

$$J(x; u(\cdot)) = \int_0^T \left(\langle QX(t), X(t) \rangle + \langle Ru(t), u(t) \rangle \right) dt + \langle GX(T), X(T) \rangle, \qquad (V.2)$$

where $Q, G \in \mathbb{S}^n$, the set of all $(n \times n)$ symmetric matrices, and $R \in \mathbb{S}^m$. The two terms on the right-hand side of the above are called the *running cost* and the *terminal cost*, respectively. Then the classical optimal control problem associated with the above linear state equation (V.1) and quadratic cost functional (V.2) can be formulated as follows.

Problem (**DLQ**)_T. For any $x \in \mathbb{R}^n$, find a $\bar{u}(\cdot) \in \mathcal{U}[0,T]$ such that

$$J(x;\bar{u}(\cdot)) = \inf_{u(\cdot)\in\mathcal{U}[0,T]} J(x;u(\cdot)).$$
(V.3)

If $\bar{u}(\cdot) \in \mathcal{U}[0,T]$ exists satisfying (V.3), we call it an *optimal control*; the corresponding $\bar{X}(\cdot) \equiv X(\cdot; x, \bar{u}(\cdot))$ and the pair $(\bar{X}(\cdot), \bar{u}(\cdot))$ are called an *optimal trajectory* and *optimal pair*, respectively. The above problem can be regarded as *linear-quadratic optimal control problem* (LQ problem, for short). The following collects the most basic conclusions for Problem (DLQ)_T.

Proposition V.1 For Problem $(DLQ)_T$, the following holds: (i) Let Problem $(DLQ)_T$ admit an optimal pair $(\bar{X}(\cdot), \bar{u}(\cdot))$. Then it is necessary that

$$R \ge 0, \tag{V.4}$$

and the cost functional $u(\cdot) \mapsto J(x; u(\cdot))$ is convex. Moreover, there exists a solution to the following ODE, called the *adjoint equation*:

$$\begin{cases} \dot{Y}(t) = -(A^{\top}Y(t) + Q\bar{X}(t)), & t \in [0,T], \\ Y(T) = G\bar{X}(T), \end{cases}$$
(V.5)

such that the following stationarity condition holds:

$$B^{\top}Y(t) + Ru(t) = 0.$$
 (V.6)

(ii) Let the optimality system:

$$\begin{cases} \bar{X}(t) = A\bar{X}(t) + B\bar{u}(t), \\ \dot{Y}(t) = -(A^{\top}Y(t) + Q\bar{X}(t)), \\ \bar{X}(0) = x, \quad Y(T) = G\bar{X}(T), \\ B^{\top}Y(t) + R\bar{u}(t) = 0. \end{cases}$$
(V.7)

admit a solution $(\bar{X}(\cdot), \bar{u}(\cdot), Y(\cdot))$, and $u(\cdot) \mapsto J(x; u(\cdot))$ be convex. Then, Problem (DLQ)_T admits an optimal control.

(iii) Let functional $u(\cdot)\mapsto J(x;u(\cdot))$ be uniformly convex, and

$$R > 0. \tag{V.8}$$

Let the following differential Riccati equation have a solution $P(\cdot)$:

$$\begin{cases} \dot{P}(t) + P(t)A + A^{\top}P(t) - P(t)BR^{-1}B^{\top}P(t) + Q = 0, \\ P(T) = G. \end{cases}$$
(V.9)

Then Problem $(DLQ)_T$ admits a unique optimal control $\bar{u}(\cdot)$, and it has the following closed-loop representation:

$$\bar{u}(t) = -R^{-1}B^{\top}P(t)\bar{X}(t), \qquad t \in [0, T].$$
 (V.10)

The above will be the case if the following canonical condition holds:

$$Q, G \ge 0, \quad R > 0. \tag{V.11}$$

Study of the LQ problems for ODEs began with the seminal works of Bellman–Glicksberg–Gross [22], Kalman [50] and Letov [57] appeared around 1960. Standard results for LQ theory of ODEs, including the above, can be found in Lee–Markus [55], Anderson–Moore [1], Willems [109], Wonham [113], and others. See also Yong–Zhou [122].

From the above results, we see that associated with the LQ problem, there are several notions closely related: Existence (and uniqueness) of optimal control, (uniform) convexity of the cost functional, optimality system (two-point boundary value problem), Riccati equation, closed-loop representation. It seems that they are almost equivalent somehow. It is a desire to make these relations clear.

On the other hand, study of stochastic LQ problems began with the works of Kushner [54] and Wonham [112] in the 1960s. See also Davis [35], Bensoussan [24], and others for classical stochastic LQ theory. In 1998, Chen-Li-Zhou [33] found that for stochastic LQ problems, the weighting matrices in the cost functional could be indefinite to some extent; in particular, (V.4) is not necessary for the existence of optimal control for the corresponding stochastic LQ problem. See Yong-Zhou [122] for some presentation of the updated theory by the end of the last century. Since the early 2010s, the authors of this paper, together with their collaborators, started to investigate the LQ problems for stochastic differential equations from a different angle. Notions of finiteness, open-loop and closed-loop solvability were introduced, together with the relationship among them and to the other relevant notions, as well as their characterizations. More precisely, the following have been established for stochastic LQ problems in finite time-horizons, denoted by Problem $(SLQ)_T$:

- If Problem (SLQ)_T is finite, then the cost functional is convex in the control process.
- The closed-loop solvability implies the open-loop solvability, but not vice versa in general.
- The open-loop solvability is equivalent to the solvability of the optimality system, which is now a forwardbackward stochastic differential equation (FBSDE, for short), plus the convexity of the cost functional.
- The closed-loop solvability is equivalent the regular solvability of a differential Riccati equation, which implicitly implies that the cost functional is convex. In this case, the problem is also open-loop solvable and any open-loop optimal control admits a closed-loop representation (or called state-feedback representation), which must be an outcome of the closed-loop optimal strategy.
- If the cost functional is uniformly convex, then Problem (SLQ)_T is uniquely closed-loop solvable, and thus also uniquely open-loop solvable.

For LQ problems in the infinite time-horizon, denoted by Problem $(SLQ)_{\infty}$, under the stabilizability condition, the square integrability of the nonhomogeneous terms and linear weight processes, the following are established:

- If Problem $(SLQ)_{\infty}$ is finite, then the cost functional is convex in the control process.
- When the cost functional is convex, the following are equivalent:
 - Problem $(SL)_{\infty}$ is open-loop solvable;
 - Problem $(SLQ)_i$ is closed-loop solvable;
 - An FBSDE over $[0,\infty)$ admits an L^2 -stable adapted solution;
 - An algebraic Riccati equation admits a stabilizing solution.

See Sun-Yong [98] for a detailed comprehensive presentation of these results.

VI. CONCLUDING REMARKS

In this survey paper, we have studied several different formulations on nonstandard LQ decision-making problems, and discussed the underlying challenges and some of their resolutions. This is clearly not a comprehensive review. There are other numerous formulations, approaches, and results on LQ problems, both of theoretical nature and those that arise in applications. The readers are directed to the references cited for other variations and applications of the rich LQ framework.

One recent research direction of LQG control is to connect it with *learning*. The main concern there is to learn model and cost parameters from environment and find efficient learning models and rates, similar to those in reinforcement learning. Some representative results in this direction can be found in [2], [29], [87], [104], [105], [132], and the references therein.

REFERENCES

- [1] B. D. O. Anderson and J. B. Moore, *Linear Optimal Control*. New Jersey: Prentice-Hall Englewood Cliffs, 1971.
- [2] D. Baby and Y.-X. Wang, "Optimal dynamic regret in LQR control," in Proceedings of the 2022 Conference on Neural Information Processing Systems, 2022.
- [3] T. Başar, "Stochastic differential games and intricacy of information structures," in J. Haunschmied, V.M. Veliov, and S. Wrzaczek (Eds), *Dynamic Games in Economics*, Springer, 2014, pp. 23-49.
- [4] —, "On the saddle-point solution of a class of stochastic differential games," *Journal on Optimization Theory and Applications*, vol. 33, no. 4, pp. 539–556, 1981.
- [5] —, "The Gaussian test channel with an intelligent jammer," *IEEE Transactions on Information Theory*, vol. 29, no. 1, pp. 152–157, 1983.
- [6] —, "Affine incentive schemes for stochastic systems with dynamic information," *SIAM Journal on Control and Optimization*, vol. 22, no. 2, pp. 199–210, 1984.
- [7] —, "An equilibrium theory for multi-person decision making with multiple probabilistic models," *IEEE Transactions on Automatic Control*, vol. 30, no. 2, pp. 118–132, 1985.
- [8] —, "Solutions to a class of nonstandard stochastic control problems with active learning," *IEEE Transactions on Automatic Control*, vol. 33, no. 12, pp. 1122–1129, 1988.
- [9] —, Dynamic optimization of some forward-looking stochastic models. Springer-Verlag, 1989, pp. 315–336, modeling and Control of Systems, volume 121 of Lecture Notes in Control and Information Sciences.
- [10] —, "Some thoughts on rational expectations models, and alternate formulations," *Computer and Mathematics with Applications*, vol. 18, no. 6/7, pp. 591–604, 1989.
- [11] —, "Stochastic incentive problems with partial dynamic information and multiple levels of hierarchy," *European J. Political Economy* (special issue on "Economic Design"), pp. 203–217, 1989.
- [12] —, Edi., Control Theory: Twenty-Five Seminal Papers. IEEE Press, 2001.
- [13] —, "Variations on the theme of the Witsenhausen counterexample," in Proc. 47th IEEE Conference on Decision and Control (CDC'08, Dec 9-11, 2008; Cancun, Mexico), 2008, pp. 1614–1619.
- [14] T. Başar and R. Bansal, "Optimum design of measurement channels and control policies for linear-quadratic stochastic systems," *European Journal of Operations Research*, vol. 73, pp. 226–236, 1994.
- [15] T. Başar and P. Bernhard, H[∞] Optimal Control and Related Minimax Design Problems, 2nd ed. Boston, MA: Birkhäuser, 1995.
- [16] T. Başar and G. J. Olsder, Dynamic Noncooperative Game Theory, 2nd ed. SIAM, 1999.
- [17] A. Bagchi and T. Başar, "Team decision theory for linear continuoustime systems," *IEEE Transactions on Automatic Control*, vol. 27, no. 6, pp. 1154–1161, 1980.
- [18] —, "Stackelberg strategies in linear-quadratic stochastic differential games," *Journal of Optimization Theory and Applications*, vol. 35, no. 3, pp. 443–464, 1981.
- [19] R. Bansal and T. Başar, "Solutions to a class of linear-quadraticgaussian lqg stochastic team problems with nonclassical information," *Systems and Control Letters*, vol. 9, no. 2, pp. 125–130, 1987.
- [20] —, "Stochastic team problems with nonclassical information revisited - when is an affine law optimal?" *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 554–559, 1987.
- [21] —, "Simultaneous design of measurement and control strategies in stochastic systems with feedback," *Automatica*, vol. 45, no. 5, pp. 679–694, 1989.
- [22] R. Bellman and I. Glicksberg, Some Aspects of the Mathematical Theory of Control Processes. Santa Monica, California: Rand Corporation, 1958.
- [23] R. Bellman, R. Kalaba, and G. M. Wing, "Invariant imbedding and the reduction of two-point boundary value problems to initial value problems," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 46, pp. 1646–1649, 1960.
- [24] A. Bensoussan, *Lectures on Stochastic Control.* Berlin: Springer-Verlag, 1981, lecture Notes in Math. Vol. 972.
- [25] A. Bensoussan, S. Chen, and S. P. Sethi, "The maximum principle for global solutions of stochastic Stackelberg differential games," *SIAM J. Control Optim.*, vol. 53, no. 4, pp. 1956–1981, 2015.
- [26] A. Bensoussan, K. C. J. Sung, S. C. P. Yam, and C. P. Yung, "Linearquadratic mean field games," 2014, https://arxiv.org/abs/1404.5741.

- [27] J. M. Bismut, "Linear quadratic optimal stochastic control with random coefficient," *SIAM Journal on Control*, vol. 14, no. 3, pp. 419–444, 1976.
- [28] J. F. Bonnans and F. J. Silva, "First and second order necessary conditions for stochastic optimal control problems," *Applied Mathematics* & *Optimization*, vol. 65, pp. 403–439, 2012.
- [29] A. Cassel and T. Karen, "Online policy gradient for model free learning of linear quadratic regulators with \sqrt{T} regret," in *Proceedings* of the 38th International Conference on Machine Learning, 2021.
- [30] D. Castanon and M. Athans, "On stochastic dynamic Stackelberg strategies," *Automatica*, vol. 12, no. 2, pp. 177–183, 1976.
- [31] T. Chan, "Dynamics of the McKean-Vlasov equation," Annals of Probability, vol. 22, pp. 431–441, 1994.
- [32] H.-F. Chen, "Unified controls applicable to general case under quadratic index," *Acta Mathematicae Applicatae Sinica*, vol. 5, pp. 45–52, 1982.
- [33] S. Chen, X. Li, and X. Y. Zhou, "Stochastic linear quadratic regulators with indefinite control weight costs," *SIAM J. Control Optim.*, vol. 36, pp. 1695–1702, 1998.
- [34] S. Chen and X. Y. Zhou, "Stochastic linear quadratic regulator with indefinite control weight costs II," *SIAM Journal on Control and Optimization*, vol. 39, no. 4, pp. 1065–1081, 2000.
- [35] M. H. A. Davis, *Linear Estimation and Stochastic Control*. London: Chapman & Hall, 1977.
- [36] R. Elliott, X. Li, and Y. H. Ni, "Discrete-time mean field stochastic linear-quadratic optimal control problems," *Automatica*, vol. 49, pp. 3222–3233, 2013.
- [37] J. Engwerda, LQ Dynamic Optimization and Differential Games. John Wiley & Sons, 2005.
- [38] W. Fleming and H. M. Soner, Controlled Markov Processes and Viscosity Solutions, 2nd ed. Springer, 2006.
- [39] W. H. Fleming and P. E. Souganidis, "On the existence of value functions of two-player zero-sum stochastic differential games," *Indiana Univ. Math. J*, vol. 38, pp. 293–314, 1989.
- [40] M. Fuhrman and G. Tessitore, "Infinite horizon backward stochastic differential equations and elliptic equations in Hilbert spaces," *The Annals of Probability*, vol. 32, no. 1B, pp. 607–660, 2004.
- [41] R. Gabasov and F. M. Kirillova, "High order necessary conditions for optimality," SIAM Journal on Control, vol. 10, pp. 127–168, 1972.
- [42] V. Gurman, "The method of multiple maxima and optimization problems for space maneuvers," in *Proc. Second Readings of K. E. Tsiolkovskii, Moscow*, 1968, pp. 39–51.
- [43] Y. Ho, "Linear stochastic singular control problems," Journal of Optimization Theory and Application, vol. 9, pp. 24–31, 1972.
- [44] Y. C. Ho and K. C. Chu, "Team decision theory and information structures in optimal control problems-part 1," *IEEE Transactions* on Automatic Control, vol. 17, no. 1, pp. 15–22, 1972.
- [45] Y. Hu and G. Tessitore, "BSDE on an infinite horizon and elliptic PDEs in infinite dimension," *Nonlinear Differential Equations and Applications*, vol. 14, no. 5, pp. 825–846, 2007.
- [46] J. Huang, X. Li, and J. Yong, "A linear-quadratic optimal control problem for mean-field stochastic differential equations in infinite horizon," *Math. Control Relat. Fields*, vol. 5, pp. 97–139, 2015.
- [47] M. Huang, P. E. Caines, and R. P. Malhame, "Uplink power adjustment in wireless communication systems: A stochastic control analysis," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1693–1708, 2004.
- [48] O. C. Imer, S. Yüksel, and T. Başar, "Optimal control of LTI systems over unreliable communication links," *Automatica*, vol. 42, no. 9, pp. 1429–1440, 2006.
- [49] M. Kac, "Foundation of kinetic theory," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, 1956, pp. 171–197.
- [50] R. E. Kalman, "Contributions to the theory of optimal control," *Bol. Soc. Mat. Mexicana*, vol. 5, pp. 102–119, 1960.
- [51] A. C. Kizilkale and R. P. Malhame, "Collective target tracking mean field control for Markovian jump-driven models of electric water heating loads," in *Proceedings of the 19th World Congress, The International Federation of Automatic Control*, 2014, pp. 1867–1872.
- [52] M. Kohlmann and S. Tang, "Multidimensional backward stochastic Riccati equations and applications," *SIAM J. Control Optim.*, vol. 41, pp. 1696–1721, 2003.
- [53] A. J. Krener, "The high order maximal principle and its application to singular extremals," *SIAM Journal of Control and Optimization*, vol. 15, pp. 256–293, 1977.

- [54] H. J. Kushner, "Optimal stochastic control," *IRE Trans. Auto. Control*, vol. 7, pp. 120–122, 1962.
- [55] E. B. Lee and L. Markus, Foundations of Optimal Control Theory. New York: John Wiley, 1967.
- [56] L. Lessard and A. Nayyar, "Structural results and explicit solution for two-player LQG systems on a finite time horizon," in *Proceedings* of 52nd IEEE Conference on Decision and Control, 2013, pp. 6542– 6549.
- [57] A. M. Letov, "Analytic design of regulators, Automat. i Telemekh," 436–441; 561–568; 661–665; 22 (1961), 425–435; 23 (1962), 1405– 1413 (in Russian).
- [58] L. Li and H. Zhang, "Stabilization of discrete-time systems with multiplicative noise and multiple delays in the control variable," *SIAM Journal on Control and Optimization*, vol. 54, no. 2, pp. 894– 917, 2016.
- [59] P. Li, A. E. B. Lim, and G. Shanthikumar, "Decentralized control of a stochastic multi-agent queueing system," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2762–2777, 2012.
- [60] X. Li, J. Shi, and J. Yong, "Mean-field linear-quadratic stochastic differential games in an infinite horizon," *ESAIM: COCV*, vol. 27, p. 81, 2021.
- [61] X. Li, J. Sun, and J. Yong, "Mean-field stochastic linear quadratic optimal control problems: closed-loop solvability," *Probab. Uncertainty* & *Quantitative Risk*, vol. 1, pp. 1–22, 2016.
- [62] X. Liang, Q. Qing, H. Zhang, and L. Xie, "Decentralized control for networked control systems with asymmetric information," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 2076–2083, 2021.
- [63] Q. Lü, T. Wang, and X. Zhang, "Characterization of optimal feedback for stochastic linear quadratic control problems," *Probab. Uncertainty, Quan. Risk*, vol. 2, p. 2, 2017.
- [64] J. Ma, P. Protter, and J. Yong, "Solving forward-backward stochastic differential equations explicitly — a four-step scheme," *Probability Theory & Related Fields*, vol. 98, pp. 339–359, 1994.
- [65] J. Ma and J. Yong, Forward-Backward Stochastic Differential Equations and Their Applications. Berlin: Springer-Verlag, 1999.
- [66] T. Ma, J. Xu, and H. Zhang, "Explicit solution to forward and backward stochastic differential equations with state delay," *Control Theory and Technology*, vol. 20, p. 303–315, 2022.
- [67] J. Moon, "Necessary and sufficient conditions of risk-sensitive optimal control and differential games for stochastic differential delayed equations," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 14, pp. 4812–4827, 2019.
- [68] —, "A sufficient condition for linear-quadratic stochastic zero-sum differential games for Markov jump systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1619–1626, 2019.
- [69] —, "A feedback Nash equilibrium for affine-quadratic zero-sum stochastic differential games with random coefficients," *IEEE Control Systems Letters*, vol. 4, no. 4, pp. 868–873, 2020.
- [70] —, "Linear-quadratic mean field stochastic zero-sum differential games," Automatica, vol. 120 (109067), pp. 1–10, 2020.
- [71] —, "The risk-sensitive maximum principle for controlled forwardbackward stochastic differential equations," *Automatica*, vol. 120 (109069), pp. 1–14, 2020.
- [72] —, "Linear-quadratic stochastic Stackelberg differential games for jump-diffusion systems," *SIAM Journal on Control and Optimization*, vol. 39, no. 2, pp. 954–976, 2021.
- [73] ——, "Backward reachability approach to state-constrained stochastic optimal control problem for jump-diffusion models," *Advances in Continuous and Discrete Model*, vol. 68, pp. 1–38, 2022.
- [74] ——, "Linear-quadratic mean-field type Stackelberg differential games for stochastic jump-diffusion systems," *Mathematical Control* and Related Fields, vol. 12, no. 2, pp. 371–404, 2022.
- [75] ——, "State and control path-dependent stochastic optimal control with jumps," *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4555–4567, 2022.
- [76] —, "Stochastic optimal control with random coefficients and associated stochastic Hamilton-Jacobi-Bellman equations," Advances in Continuous and Discrete Models, vol. 3, pp. 1–32, 2022.
- [77] —, "Linear-quadratic stochastic leader-follower differential games for Markov jump-diffusion models," *Automatica*, vol. 147 (110713), pp. 1–18, 2023.
- [78] J. Moon and T. Başar, "Minimax control over unreliable communication channels," *Automatica*, vol. 59, pp. 182–193, 2015.

- [79] —, "Minimax estimation with intermittent observations," *Automatica*, vol. 62, pp. 122–133, 2015.
- [80] —, "Linear quadratic mean field Stackelberg differential games," *Automatica*, vol. 97, pp. 200–213, 2018.
- [81] —, "Dynamic programming and verification theorem of recursive stochastic control problem for jump-diffusion models with random coefficients," *IEEE Transactions on Automatic Control*, vol. 67, no. 12, pp. 6474–6488, 2022.
- [82] J. Moon and Y. Kim, "Linear-exponential-quadratic control for mean field stochastic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 5094–5100, 2019.
- [83] J. Moon and H. J. Yang, "Linear-quadratic time-inconsistent meanfield type Stackelberg differential games: Time-consistent open-loop solutions," *IEEE Transactions on Automatic Control*, vol. 66, no. 1, pp. 375–382, 2021.
- [84] J. Moore, "The singular solutions to a singular quadratic minimization problem," *International Journal of Control*, vol. 20, pp. 383–393, 1974.
- [85] L. Mou and J. Yong, "Two-person zero-sum linear quadratic stochastic differential games by a Hilbert space method," *Industrial & Management Optim.*, vol. 2, pp. 95–117, 2006.
- [86] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, p. 1644–1658, 2013.
- [87] Y. Park, R. Rossi, Z. Wen, G. Wu, and H. Zhao, "Structured policy iteration for linear quadratic regulator," in *Proceedings of the 37th International Conference on Machine Learning*, 2020.
- [88] M. A. Rami and X. Y. Zhou, "Linear matrix inequalities, Riccati equations, and indefinite stochastic linear quadratic controls," *IEEE Transactions on Automatic Control*, vol. 45, no. 6, pp. 1131–1143, 2000.
- [89] D. W. Ross, "Controller design for time lag systems via a quadratic criterion," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 664–672, 1971.
- [90] M. Scheutzow, "Uniqueness and non-uniqueness of solutions of Vlasov-McKean equations," *Journal of the Australian Mathematical Society*, vol. 43, no. 2, pp. 246–256, 1987.
- [91] O. J. Smith, "A controller to overcome dead time," *ISA Journal*, vol. 6, no. 2, pp. 28–33, 1959.
- [92] J. Sun, X. Li, and J. Yong, "Open-loop and closed-loop solvabilities for stochastic linear quadratic optimal control problems," *SIAM Journal on Control and Optimization*, vol. 54, pp. 2274–2308, 2016.
- [93] J. Sun, J. Xiong, and J. Yong, "Indefinite stochastic linear-quadratic optimal control problems with random coefficients: Closed-loop representation of open-loop optimal controls," *Annals of Applied Probability*, vol. 31, pp. 460–499, 2021.
- [94] J. Sun and J. Yong, "Linear quadratic stochastic differential games: open-loop and closed-loop saddle points," *SIAM J. Control & Optim.*, vol. 52, pp. 4082–4121, 2014.
- [95] —, "Stochastic linear quadratic optimal control problems in infinite horizon," *Appl. Math. Optim.*, vol. 78, pp. 145–183, 2018.
- [96] —, "Linear quadratic stochastic two-person nonzero-sum differential games: open-loop and closed-loop Nash equilibria," *Stoch. Proc. Appl.*, vol. 129, pp. 381–418, 2019.
- [97] —, Stochastic Linear-Quadratic Optimal Control Theory: Differential Games and Mean-Field Problems. Springer, 2020, Springer Brief in Math.
- [98] —, Stochastic Linear-Quadratic Optimal Control Theory: Open-Loop and Closed-Loop Solutions. Springer, 2020, Springer Brief in Math.
- [99] J. Sun, J. Yong, and S. Zhang, "Linear quadratic stochastic twoperson zero-sum differential games in an infinite horizon," *ESAIM: COCV*, vol. 22, pp. 743–769, 2016.
- [100] S. Tang, "General linear quadratic optimal stochastic control problems with random coefficients: linear stochastic Hamilton systems and backward stochastic Riccati equations," *SIAM J. Control Optim.*, vol. 42, pp. 53–75, 2003.
- [101] —, "Dynamic programming for general linear quadratic optimal stochastic control problems with random coefficients," *SIAM J. Control Optim.*, vol. 53, pp. 1082–1106, 2015.
- [102] H. P. McKean, "A class of Markov processes associated with nonlinear parabolic equations," *Proceedings of National Academy of Science (PNAS) of the United States of America*, vol. 56, pp. 1907– 1911, 1966.

- [103] N. Touzi, Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE. Springer, 2013.
- [104] M. A. uz Zaman, E. Miehling, and T. Başar, "Reinforcement learning for non-stationary discrete-time linear-quadratic mean-field games in multiple populations," *Dynamic Games and Applications*, vol. 1, no. 13, pp. 118–164, 2023.
- [105] M. A. uz Zaman, K. Zhang, E. Miehling, and T. Başar, "Reinforcement learning in non-stationary discrete-time linear-quadratic mean-field games," in *Proceedings of the 59th IEEE Conference on Decision and Control*, 2020, pp. 2278–2284.
- [106] H. Wang, H. Zhang, and L. Xie, "Optimal control and stabilization for Itô systems with input delay," *Journal of System Science & Complexity*, vol. 34, no. 5, pp. 1895–1926, 2021.
- [107] V. M. Warfield, "A stochastic maximum principle," SIAM Journal on Control, vol. 14, no. 5, pp. 803–826, 1976.
- [108] Q. Wei, J. Yong, and Z. Yu, "Linear quadratic optimal control problems with operator coefficients: open-loop solutions," *ESAIM COCV*, vol. 25, p. 17, 2019.
- [109] J. C. Willems, "Least squares stationary optimal control and the algebraic Riccati equation," *IEEE Trans. Automat. Control*, vol. 16, pp. 621–634, 1971.
- [110] —, "Singular optimal control: a geometric approach," IAM Journal of Control and Optimization, vol. 24, pp. 323–337, 1986.
- [111] H. S. Witsenhausen, "A counter example in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [112] W. M. Wonham, "On a matrix Riccati equation of stochastic control," SIAM J. Control, vol. 6, pp. 681–697, 1968.
- [113] —, Linear Multivariable Control: A Geometric Approach, 2nd ed. Springer-Verlag, 1979.
- [114] J. Xu, W. Wang, and H. Zhang, "Solution to delayed forward and backward stochastic difference equations and its applications," *IEEE Transactions on Automatic Control*, vol. 66, no. 3, pp. 1407–1413, 2021.
- [115] J. Xu and H. Zhang, "Output feedback control for irregular LQ problem," *IEEE Control Systems Letters*, vol. 5, no. 3, pp. 875–880, 2021.
- [116] J. Xu, H. Zhang, and T. Başar, "Stackelberg solution for a two-agent rational expectations model," *Automatica*, vol. 129, no. 109601, pp. 1–8, 2021.
- [117] —, "Decentralized LQG control with d-step delayed information sharing pattern," *IEEE Transactions on Automatic Control*, vol. 68, no. 1, pp. 604–611, 2023.
- [118] J. Xu, H. Zhang, and L. Xie, "General linear forward and backward stochastic difference equations with applications," *Automatica*, vol. 96, pp. 40–50, 2018.
- [119] J. Yong, "A linear-quadratic optimal control problem for mean-field stochastic differential equations," *SIAM J. Control*, vol. 51, pp. 2809– 2838, 2013.
- [120] —, Differential Games: A Concise Introduction. World Scientific, 2015.
- [121] —, "Linear-quadratic optimal control problems for mean-field stochastic differential equations-time-consistent solutions," *Transactions on the American Mathematical Society*, vol. 369, no. 8, pp. 5467–5523, 2017.
- [122] J. Yong and X. Y. Zhou, Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer, 1999.
- [123] J. Yong, "A leader-follower stochastic linear quadratic differential game," SIAM J. Control Optim., vol. 41, no. 4, pp. 1015–1041, 2002.
- [124] Z. Yu, "An optimal feedback control-strategy pair for zero-sum linearquadratic stochastic differential game: the Riccati equation approach," *SIAM Journal on Control and Optimization*, vol. 53, no. 4, pp. 2141– 2167, 2015.
- [125] S. Yüksel and T. Başar, "Minimum rate coding for LTI systems over noiseless channels," *IEEE Transactions on Automatic Control*, vol. 51, no. 12, pp. 1878–1887, 2006.
- [126] —, "Control over noisy forward and reverse channels," *IEEE Transactions on Automatic Control*, vol. 56, no. 5, pp. 1014–1029, 2011.
- [127] —, Stochastic Networked Control Systems. Springer, 2013.
- [128] H. Zhang and J. Xu, "Control for Itô stochastic systems with input delay," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 350–365, 2017.
- [129] —, "Optimal control with irregular performance," Science China Information Sciences, vol. 62, no. 192203, pp. 1–14, 2019.

- [130] —, "The difference and unity of irregular LQ control and standard LQ control and its solution," *ESAIM: COCV*, vol. 27, no. S31, 2021.
- [131] H. Zhang and X. Zhang, "Pointwise second-order necessary conditions for stochastic optimal controls, part 1: The case of convex control constraint," *SIAM Journal of Control and Optimization*, vol. 53, no. 4, pp. 2267–2296, 2015.
- [132] K. Zhang, Z. Yang, and T. Başar, "Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games," in *Proceedings of the 2019 Conference on Neural Information Processing Systems*, 2019.
- [133] Z. Y. Zhou and D. Li, "Continuous-time mean-variance portfolio selection: A stochastic LQ framework," *Applied Mathematics and Optimization*, vol. 42, pp. 19–33, 2000.