Deterministic Optimality of the Steady-State Behavior of the Kalman-Bucy Filter

Deterministic optimality of the steady-state behavior of the Kalman-Bucy filter

Corrado Possieri and Mario Sassano

***Abstract*—In this paper, we provide a deterministic characterization of optimality of the steady-state behavior of the Kalman-Bucy filter, via an inverse optimal control argument. The result is achieved in two steps, both interesting *per se*. First, a singular linear-quadratic (LQ) optimal control problem is formulated and solved with respect to the innovation term of a classic Luenberger observer, hence yielding a LQ optimal observer. Then, such a construction is employed to interpret the optimality of the steady-state behavior of the celebrated Kalman-Bucy filter in a purely deterministic sense.**

***Index Terms*—Observers for linear systems, optimal control, Kalman filtering.**

I. Introduction

SINCE its seminal formalizations [1], [2], [3], the problem of reconstructing (part of) the state of a dynamical system that is not directly measurable has acquired a role of paramount importance in systems and control theory [4], [5], [6], [7], [8], [9], [10], [11]. Two, somewhat alternative, approaches to the observer design task have immediately emerged. On one hand, those based on a perfect knowledge of the dynamics of a completely deterministic plant with the aim of designing an auxiliary systems (the *observer*) - formally belonging to the same class of the original plant - whose state trajectories converge asymptotically to the state trajectories of the observed system. On the other hand, those based on the premise that the observed plant and measurements are corrupted by random processeses (noise), with certain (typically known) stochastic properties with the aim of designing an auxiliary system (usually referred to as a *filter*) such that the expected squared error is minimized. To further substantiate the above dichotomy, the most celebrated candidates of each class are the so-called *Luenberger observer* [2] and the *Kalman-Bucy filter* [3], respectively.

An additional difference between the alternative strategies consists in the fact that, typically, the design of the former object (Luenberger observer) is not driven by any optimality considerations as long as the correction term of the auxiliary system is such that the trajectories converge to the original ones, while the latter (Kalman-Bucy filter) are motivated by stochastic optimality considerations, provided the noise belong to specially structured classes of random processes. Nonetheless, it is interesting to point out that, despite such seemingly opposite *philosophies* behind their construction, the two underlying auxiliary systems possess rather similar structures. Moreover, in order to put the contribution of this paper into the correct perspective, it is worth stressing that, despite the fact that the latter results have been explored also in a deterministic setting via the comprehensive *duality* theory [12], [13] or deterministic optimal filtering [14], these interpretations are still fundamentally different from the ones discussed here. In fact, the aim of, *e.g.*, [14] consists in formulating the filtering problem as a *virtual* optimal control problem in which one is essentially allowed to *select* the deterministic disturbances in such a way that the resulting trajectory is *consistent* with the measured output and the $\mathcal{L}_2$-norm of the disturbances is (virtually) minimized. Here, instead, the focus is on minimizing the *innovation term* together with the transient response of the observer in the absence of (stochastic) noise or (deterministic) disturbances. Such considerations on the magnitude of the innovation term may be particularly beneficial in practical applications with the objective of avoiding excessively demanding control actions in a physical plant in closed loop with a (digital) observer. Indeed, the state of the observer is typically fed back to the system via a stabilizing gain, hence diminishing the magnitude of the innovation term may be beneficial for reducing the energy of the control input applied in the closed loop. Alternatively, one may envision using saturation functions in the overall control scheme to compensate for the transient behavior of the observer, with the detrimental effect, however, of introducing unnecessary nonlinearities in the closed-loop system. This aspect has not been pursued hitherto in the literature, to the best of our knowledge, due to the arising challenge in the design process to tackle the *feasibility* requirements of expressing the optimal solution only in terms of available information on the state.

The main objective of this note consists in addressing simultaneously the two aspects identified in the above discussion. First, we state necessary and sufficient conditions that allow to design a classic Luenberger observer according to a given quadratic cost functional. Then, we exploit this result to provide a *deterministic characterization* of the optimality of the steady-state behavior of the Kalman-Bucy filter.

Notation. Given $M \in \mathbb{R}^{n\times n}$, $M^\dagger$ denotes the Moore-Penrose pseudoinverse of M, whereas $\mathrm{im}(M)$ and $\ker(M)$ denote the image and the kernel of M, respectively. Provided M is symmetric, the notation $M \succ 0$ ($M \succeq 0$) specifies that M is positive definite (positive semi-definite). The symbol $\mathrm{vec}(M)$ denotes the vectorization of M, whereas $A \otimes B$ denotes the Kronecker product within A and B.

II. Problem statement

Consider continuous-time linear systems described by

$$\dot{x} = Ax + Bu, \qquad y = Cx, \tag{1}$$

C. Possieri is with Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, 10129 Torino, Italy.

M. Sassano is with Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università di Roma Tor Vergata, Roma, 00133, Italy.

Email: corrado.possieri@polito.it, mario.sassano@uniroma2.it

 DOI: 10.1109/LCSYS.2019.2918488

where $x(t) \in \mathbb{R}^n$ denotes the state, $u(t) \in \mathbb{R}^m$ the input, and $y(t) \in \mathbb{R}^p$ the measured output. Assuming that the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$ are precisely known and that the input signal u can be measured, an observer for (1) consists of a dynamical system of the form

$$\dot{\hat{x}} = A\hat{x} + Bu + v, \tag{2}$$

where $\hat{x}(t) \in \mathbb{R}^n$ yields an estimate of the state $x(t)$ of system (1) and $v \in \mathbb{R}^n$ is the *correction input* that, on the basis of measurements of the output y, is used to steer the state $\hat{x}$ of system (2) to the state x of system (1). Since the choice of the correction term v is not unique, it may be reasonable to introduce an optimality criterion to induce an *ordering* of such viable correction inputs. To this end, define the error $e := \hat{x} - x$, whose dynamics are governed by

$$\dot{e} = Ae + v, \tag{3}$$

and consider the cost functional

$$J(e_0, v) = \int_0^\infty \left(e^\top(t)Qe(t) + v^\top(t)Rv(t)\right) \mathrm{d}t, \tag{4}$$

with $Q = Q^\top \in \mathbb{R}^{n\times n}$, $Q \succeq 0$, $R = R^\top \in \mathbb{R}^{n\times n}$, $R \succeq 0$, and $e_0 := e(0)$. Note that, since R is required to be positive semi-definite, *singular* problems are potentially allowed (see [15]): the rationale behind this choice will become evident in the following sections while establishing the connections with deterministic optimality of the Kalman-Bucy filter. Note that the cost functional (4) can be envisioned as a generic integral quadratic form in e and $\dot{e}$. Within the framework defined above, the main objective of this paper is twofold. First, a continuous correction input $v^\star(t)$ that depends only on the output mismatch $\hat{y} - y = C(\hat{x} - x) = Ce$ is determined to minimize the cost functional $J(e_0, v)$ given in (4) subject to the dynamics given in (3) for each initial error $e_0 \in \mathbb{R}^n$. Then, it is shown that such result is instrumental for establishing a deterministic characterization - via an *inverse optimal control* argument - of optimality of the celebrated Kalman-Bucy filter. Note that establishing such a relation is different from the results derived in the celebrated duality theory, since (1) is not affected by any (stochastic or deterministic) disturbance.

III. Linear quadratic optimal observers

The objective of this section is to characterize the optimal correction input $v^\star$ mentioned in the previous section. In particular, considering a general *singular* Linear Quadratic (LQ) control problem and provided Q is positive semidefinite, it can be claimed, by relying on [15, Lem. 2] and on [16, Thm. 2], that if there exists an optimal solution to the singular LQ problem described by (3) and (4) for any initial condition $e_0 \in \mathbb{R}^n$, then such an optimal solution can be written in the form of linear state feedback as

$$v^* = \tilde{K}^* e, \tag{5}$$

for some $\tilde{K}^* \in \mathbb{R}^{n\times n}$. However, in addition to the constraints induced by the singular formulation of the LQ control problem, as in [15], [16], herein the further requirement of designing an optimal correction input that depends only on the measurable *output mismatch* Ce must be considered. Therefore, the following statement formalizes the concept of *admissible* linear, feedback correction input v, which, by the reasoning given above, must be in the form of a linear feedback of the error mismatch.

Definition 1 (Admissible correction input)**.** A correction input $v(t)$ is *admissible* if there exists $K \in \mathbb{R}^{n\times p}$ such that

$$v(t) = KCe(t)\,, \tag{6}$$

for any $t \geqslant 0$. ◦

It is worth pointing out that, by Definition 1, each admissible correction input v is such that the corresponding dynamics (2) indeed constitutes a Luenberger observer [1], [2] for system (1), provided that the matrix $A + KC$ is Hurwitz. The following theorem, the proof of which is postponted to the Appendix A, provides necessary and sufficient conditions for the existence of an *optimal admissible correction input*.

Theorem 1. *Consider the error dynamics* (3) *and the cost functional* (4)*. There exists an* admissible optimal correction input $v^\star$ *if and only if there is* $P = P^\top \in \mathbb{R}^{n\times n}$*,* $P \succeq 0$*, that solves the* generalized Riccati equality

$$PA + A^\top P + Q - PR^\dagger P = 0 \tag{7a}$$

with the additional constraint

$$\mathrm{vec}(P) \in \mathrm{im}\left(C^\top \otimes R\right). \tag{7b}$$

In such a case, letting $P^\star$ *be the smallest of such solutions*[1]*, an optimal* $K^\star$ *can be obtained by solving*

$$RK^\star C = -P^\star, \tag{8}$$

and the corresponding value of the cost functional is

$$J^\star(e_0) = \min_v J(e_0, v) = e_0^\top P^\star e_0. \qquad \blacksquare$$

The following remarks provide further insights on the conditions given in (7) under some additional assumptions.

Remark 1. The equations (7) resemble the *constrained generalized continuous algebraic Riccati equation* (briefly, *CGCARE*), which arise when dealing with singular optimal control problems [15]. As anticipated above, the crucial difference between (7) herein and equations (6), (7) of [15] consists in the fact that here the control input $v^\star$ is synthesized only on the basis of the output mismatch Ce rather than relying on the availablity of the entire state variable e of system (3), which would not, in fact, lead to a feasible Luenberger's observer for (2). Nonetheless, if $C = I$, *i.e.*, the whole state of system (1) is measured, then (7) straightforwardly reduce to the CGCARE given in [15]. In fact, if $C = I$, then (7b) reduces to $\mathrm{vec}(P) \in \mathrm{im}\left(I^\top \otimes R\right)$, which holds if and only if there exists a matrix L such that $(I^\top \otimes R)\,\mathrm{vec}(L) = P$, *i.e.*, if and only if

$$R\,L = P. \tag{9}$$

Note that, by [17], there exists L such that (9) holds if and only if $\mathrm{rank}(R) = \mathrm{rank}([R,\, P])$, where $[R,\, P]$ denotes the matrix

[1] By [15], there exists a positive semidefinite solution $P^\star$ to (7a) such that $P - P^\star \succeq 0$ for all solutions $P \succeq 0$ to (7a).

obtained by considering the union of the columns of R and those of P, *i.e.*, if and only if $\ker(R) \subseteq \ker(P)$. Therefore, if $C = I$, then (7b) yields precisely the constraint given in equation (7) of [15], and overall (7) reduce to equations (6) and (7) in [15]. △

Remark 2. The constrain given in (7b) can be simplified in the case that $R \succ 0$. As a matter of fact, by further extending the reasoning employed in Remark 1, if $R \succ 0$ (and hence it is nonsingular), then (7b) holds if and only if there exists a matrix L such that $LC = R^{-1}P$, *i.e.*, equivalently,

$$C^\top L^\top = PR^{-1}. \tag{10}$$

By the same arguments used in Remark 1, there exists L such that (10) holds if and only if $\operatorname{rank}(C^\top) = \operatorname{rank}([C^\top,\ PR^{-1}])$. Therefore, provided R is positive definite and by considering that $\operatorname{im}(PR^{-1}) = \operatorname{im}(P)$, the requirement (10) can be equivalently replaced by

$$\operatorname{im}(P) \subseteq \operatorname{im}(C^\top).$$ △

Remark 3. By combining the results given in Remarks 1 and 2, it can be easily derived that if $R \succ 0$ and $C = I$, then (7b) is always satisfied. In such a case, a solution to the optimal observation problem (3), (4) is given by (5), with $\tilde{K}^* = -R^{-1}\tilde{P}^\star$, where $\tilde{P}^\star$ is the smallest positive semi-definite solution to the classical algebraic Riccati equation

$$PA + A^\top P + Q - PR^{-1}P = 0,$$

which, by classical optimal control arguments [18], always exists due to the fact that the pair (A, I) is stabilizable. Furthermore, if the pair (A, Q) is detectable, then the observer (2) obtained by substituting v with v^* in (5) is a Luenberger observer for system (1). △

Finally, in the following remark, we show how to select the gain matrix $K^\star$ so that the corresponding system (2), (6) is a state observer for system (1).

Remark 4. If there exists a positive semidefinite solution to (7), letting $P^\star$ be defined as in Theorem 1, the gain $K^\star$ can be selected by solving the quadratic matrix inequality

$$RKC = -P^\star \tag{11a}$$
$$WA^\top + AW + KW + WK^\top \prec 0, \tag{11b}$$
$$W \succ 0. \tag{11c}$$

Indeed, if there is a solution to (11), then $e^\top We$ is a Lyapunov function for the closed-loop error system. On the other hand, if (11) does not admit any solution, then there does not exist an admissible optimal correction input that makes the closed-loop error system asymptotically stable. △

The section is concluded by showing that the Riccati equation above can be employed also for *policy evaluation*, namely to assess the performance of an admissible correction input $v = \bar{K}Ce$, not necessarily optimal, such that the corresponding system (3) is asymptotically stable (*i.e.*, equivalently, the matrix $A+\bar{K}C$ is Hurwitz). Namely, by [19], given a feedback gain K such that $A+\bar{K}C$ is Hurwitz, letting $\bar{P}$ be the solution to the following Sylvester equation

$$(A+\bar{K}C)^\top \bar{P} + \bar{P}(A+\bar{K}C) + Q + C^\top \bar{K}^\top R\bar{K}C = 0, \tag{12}$$

then the corresponding value of the cost given in (4) is

$$\bar{J}(e_0) = e_0^\top \bar{P} e_0.$$

Note that since $A+\bar{K}C$ is Hurwitz by assumption, by [20], there always exist a unique solution $\bar{P}$ to (12).

IV. A deterministic interpretation of the Kalman-Bucy filter

The (stationary) Kalman-Bucy filter [3] is a dynamical observer of the form (2), in which the correction input v is

$$v^\diamond = -\Pi^\diamond C^\top W^{-1} Ce, \tag{13}$$

where $\Pi^\diamond = \Pi^{\diamond\top}$, $\Pi^\diamond \succeq 0$, is the smallest solution to the *dual algebraic Riccati equation* (briefly, *ARE*)

$$A\Pi + \Pi A^\top + V - \Pi C^\top W^{-1} C\Pi = 0, \tag{14}$$

where $V = V^\top \in \mathbb{R}^{n\times n}$, $V \succeq 0$, and $W = W^\top \in \mathbb{R}^{p\times p}$, $W \succ 0$. The following remarks recall, for completeness, the optimality of the Kalman-Bucy filter in the stochastic and in the deterministic settings.

Remark 5. Consider the linear system (which is essentially derived from system (1) assuming that both the dynamics and the output are affected by stochastic noises)

$$\dot{x} = Ax + Bu + \eta, \qquad y = Cx + w, \tag{15}$$

where $\eta(t) \in \mathbb{R}^n$ and $w(t) \in \mathbb{R}^p$ are Gaussian, zero-mean, white-noise processes, with

$$\mathbb{E}[\eta(t)] = 0, \qquad \mathbb{E}[\eta(t)\eta^\top(s)] = V\delta(t-s),$$
$$\mathbb{E}[w(t)] = 0, \qquad \mathbb{E}[w(t)w^\top(s)] = W\delta(t-s),$$

where $\delta(\cdot)$ denotes the Dirac delta and $\mathbb{E}[\cdot]$ is the expected value operator. Assume that the initial condition $x(t_0)$ of system (15) has Gaussian distribution with mean $\bar{x}_0$ and covariance matrix Π_0 and that w, η and $x(t_0)$ are uncorrelated with each other. By [21], assuming that the pair (A, C) is detectable and letting $\Pi_0 = 0$ and $t_0 \to -\infty$, one has that the Kalman-Bucy filter given by (2), (13) and initialized at $\hat{x}(t_0) = \bar{x}(t_0)$ minimizes

$$\mathbb{E}[(\hat{x}(t) - x(t))^\top (\hat{x}(t) - x(t))],$$

and the error covariance is given by

$$\mathbb{E}[(\hat{x}(t) - x(t))(\hat{x}(t) - x(t))^\top] = \Pi^\diamond.$$ △

Remark 6. The steady-state behavior of the Kalman-Bucy filter can be interpreted also in a purely deterministic setting using the principle of least square estimation [12], [14]. Indeed, consider the problem (referred to as *minimum-energy estimation*) of finding a trajectory for system (15) that is consistent with the input $u(\tau)$ and the output $y(\tau)$ (namely, such that the output generated by the model coincides with the measured one) for $\tau \in (-\infty, t]$, and such that

$$J_{MEE} = \int_{-\infty}^{t} \left(\eta^\top(\tau)\tilde{W}\eta(\tau) + w^\top(\tau)\tilde{V}w(\tau)\right) \mathrm{d}\tau$$

is minimized. In [12], [14], such a problem is solved using the *dual* ARE (14) with $W = \tilde{W}^{-1}$ and $V = \tilde{V}^{-1}$. △

It is worth noticing that the setting of Section III is rather different from the one reviewed in Remarks 5 and 6. Indeed, the main objective of Section III is to determine an innovation term $v^\star$ that minimizes the cost functional $J(e_0, v)$ given in (4), which weights both the transient error and the innovation effort, in a deterministic and noiseless setting. This motivates the fact that, in order to solve such a problem, one has to solve the CGCARE (7) rather than the dual ARE (14), to which lead both the classic (stochastic) interpretation of the Kalman-Bucy filter as well as the one suggested by the duality theory. Moreover, it is worth noticing that the dual ARE (14) and the CGCARE (7) are not equivalent under the duality relations given in [3, (16)] even in the singular case $W \succeq 0$ [13, Sec. 4.3.4]. Indeed, as highlighted in Remark 1, the CGCARE (7) generalizes the one obtained for classical singular optimal control problems when taking into account the admissibility constrain (6).

The main objective of this section is to provide a deterministic interpretation of the Kalman-Bucy filter (2), (13). Thus, consider the next result, whose proof is given in Appendix B.

Theorem 2. *Consider the system* (15) *and let V and W be given. Moreover, let $\Pi^\diamond = \Pi^{\diamond\top}$, $\Pi^\diamond \succeq 0$, be the smallest solution to* (14)*. Defining $S \triangleq C^\top W^{-1} C$, if there exist symmetric matrcies Q and R such that*

$$Q \succeq 0, \tag{16a}$$

$$R \succeq 0, \tag{16b}$$

$$R\Pi^\diamond S + S\Pi^\diamond R \succeq 0, \tag{16c}$$

$$R\Pi^\diamond S - S\Pi^\diamond R = 0, \tag{16d}$$

$$R\Pi^\diamond SA + A^\top S\Pi^\diamond R + Q - S\Pi^\diamond R\Pi^\diamond S = 0, \tag{16e}$$

then the Kalman-Bucy filter (2), (13) *solves the optimal observation problem* (3), (4) *corresponding to such matrices. On the other hand, if* (16) *does not admit a (nontrivial) solution, then there does not exist a (nontrivial) cost functional of the form* (4) *such that the Kalman-Bucy filter is optimal.* ■

By Theorem 2, if (16) does not admit a nontrivial solution, then the corresponding Kalman-Bucy filter is not optimal with respect to whatsoever integral quadratic cost functional considering the deterministic noiseless setting of this paper. The next corollary is a straightforward consequence of Theorem 1 and Theorem 2 and of the fact that the matrix $S = C^\top W^{-1} C$ is symmetric and positive semi-definite.

Corollary 1. *Consider the system* (15) *and let V and W be given. Moreover, let $\Pi^\diamond = \Pi^{\diamond\top}$, $\Pi^\diamond \succeq 0$, be the smallest solution to* (14)*, and let $S = C^\top W^{-1} C$. If*

$$S\Pi^\diamond S\Pi^\diamond S - S\Pi^\diamond SA - A^\top S\Pi^\diamond S \succeq 0, \tag{17}$$

then the Kalman-Bucy filter (2), (13) *solves the deterministic optimal observation problem* (3), (4)*, with $Q = S\Pi^\diamond S\Pi^\diamond S - S\Pi^\diamond SA - A^\top S\Pi^\diamond S$ and $R = S$. Furthermore, the value of the cost functional* (4) *is given by*

$$J^\diamond(e_0) = e_0^\top S\Pi^\diamond S e_0. \quad ■$$

Corollary 1 motivates our interest in characterizing, in Section III, correction gains that are optimal with respect to *singular* optimization problems. As a matter of fact, if $p < n$, then the matrix $S = C^\top W^{-1} C$ is singular (although positive semidefinite) and hence the choice $R = S$ leads to a singular optimization problem. Furthermore, although the condition given in Corollary 1 is just sufficient, it is simpler to verify than checking the existence of a solution to (16).

V. Example

Consider a double integrator described by

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \tag{18a}$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} x. \tag{18b}$$

with $x \in \mathbb{R}^2$, $u \in \mathbb{R}$ and $y \in \mathbb{R}$. It is assumed that the dynamics (18a) are affected by unknown disturbances while the output (18b) is corrupted by measurement noise, both described as Gaussian, zero-mean processes with covariance matrices $V = \frac{1}{\varrho^2} I$ and $W = I$, respectively, where ϱ is a nonnegative scalar parameter. The above setting may, for instance, capture the scenario in which one is interested in estimating the ground position of a mobile robot via GPS measurements affected by random noise. To this end, the steady-state behavior of the Kalman-Bucy filter is given by the observer (2), with v given by

$$v = \begin{bmatrix} \frac{\sqrt{2\varrho+1}}{\varrho} & 0 \\ \frac{1}{\varrho} & 0 \end{bmatrix} e. \tag{19}$$

It may be of interest to assess whether such choice is optimal with respect to a deterministic criterion. Thus, note that the smallest positive semidefinite solution to (14) is

$$\Pi^\diamond = \begin{bmatrix} \frac{\sqrt{2\varrho+1}}{\varrho} & \frac{1}{\varrho} \\ \frac{1}{\varrho} & \frac{\sqrt{2\varrho+1}}{\varrho^2} \end{bmatrix}.$$

Hence, letting $S = C^\top W^{-1} C$ consider the linear matrix inequality (16). Although the matrix given in (17) is not positive semidefinite for all $\varrho \in \mathbb{R}$, $\varrho > 0$, the set of all the solutions to the LMI (16) can be parametrized as

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad R = \begin{bmatrix} 1 & -\sqrt{2\varrho+1} \\ -\sqrt{2\varrho+1} & 2\varrho+1 \end{bmatrix} r, \tag{20}$$

for all nonnegative ϱ, where r is a nonnegative scalar. The definitions in (20) identify a family of cost functionals, parameterized with respect to r, for which the correction term (19) yielded by the Kalman-Bucy filter is optimal in a deterministic sense. It is worth highlighting that any cost functional for which v is optimal does pose any weight whatsoever on the transient behavior of the estimation error. Figure 1 depicts the correction input given in (19) and the behavior of the error e with $\varrho = 1$ and $e(0) = \begin{bmatrix} 1 & 3 \end{bmatrix}^\top$. Note that v is such that $Rv(t) = 0$ for all $t \in \mathbb{R}_{\geqslant 0}$.

VI. Conclusions

In this paper, we provide a characterization of the deterministic properties of the steady-state behavior of the Kalman-Bucy filter. This is achieved by borrowing an inverse optimal control argument and by relying a preliminary result, contained in the paper and interesting *per se*, that allows to formulate

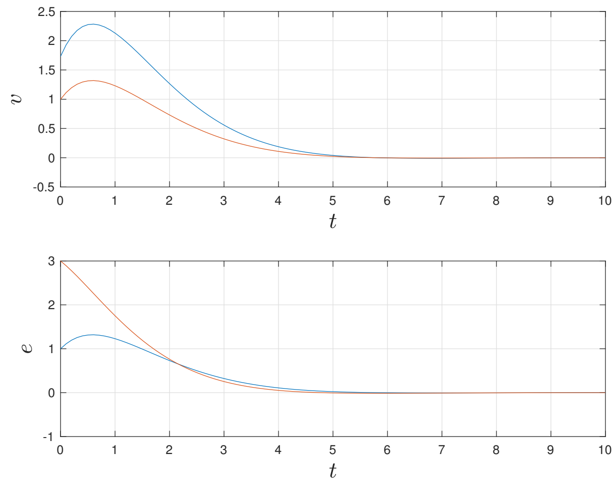

Fig. 1: Correction input and time behavior of the error.

the problem of designing the correction term of a Luenberger observer as an LQ optimal control problem with feasibility constraints on the corresponding feedback.

APPENDIX

A. Proof of Theorem 1

Two preliminary technical lemmas are discussed, which are instrumental for the proof of Theorem 1. To provide a concise statement of the following results, let $E_{j,k}^{m\times p}$ denote the matrix of dimension $m\times p$ with all the elements equal to zero except the entry of position (j,k), which equals one. Hence, define the permutation matrices $\overline{U}_{m\times p} = \sum_{j=1}^{m}\sum_{k=1}^{p} E_{j,k}^{m\times p} \otimes E_{j,k}^{m\times p}$, $U_{m\times p} = \sum_{j=1}^{m}\sum_{k=1}^{p} E_{j,k}^{m\times p} \otimes E_{k,j}^{p\times m}$ and consider the following two technical lemmas.

Lemma 1. *For any $a\in\mathbb{R}^m$ and $b\in\mathbb{R}^p$, one has*

$$(I_m\otimes a^\top)\overline{U}_{m\times p}(I_p\otimes b) = ab^\top. \quad \blacksquare$$

Proof. By relying on [22], [23], [24], it follows that

$$\begin{aligned}
&(I_m\otimes a^\top)\bar{U}_{m\times p}(I_p\otimes b)\\
&=\sum_{j=1}^{m}\sum_{k=1}^{p}(I_m\otimes a^\top)(E_{j,k}^{m\times p}\otimes E_{j,k}^{m\times p})(I_p\otimes b)\\
&=\sum_{j=1}^{m}\sum_{k=1}^{p}E_{j,k}^{m\times p}\otimes (a^\top E_{j,k}^{m\times p} b)\\
&=\sum_{j=1}^{m}\sum_{k=1}^{p}(I_m\otimes b^\top)(E_{j,k}^{m\times p}\otimes E_{k,j}^{p\times m})(I_p\otimes a)\\
&=(I_m\otimes b^\top)U_{m\times p}(I_p\otimes a) = (b^\top\otimes I_m)(I_p\otimes a) = ab^\top.
\end{aligned}$$

thus concluding the proof. $\square$

Lemma 2. *Consider two matrices $M\in\mathbb{R}^{m\times n}$ and $N\in\mathbb{R}^{n\times p}$, and suppose that $N\neq 0$. Then, it follows that $Mxx^\top N=0$ for all $x\in\mathbb{R}^n$ if and only if $M=0$.* $\blacksquare$

Proof. Sufficiency is straightforward: if M is the identically zero matrix, then $Mxx^\top N=0$ for all $x\in\mathbb{R}^n$. On the other hand, to show necessity, note that $x\in\mathbb{R}^n$ is such that

$$Mxx^\top N = 0 \tag{21}$$

if and only if one of the following two statements hold:

(i) $Mx=0$, which implies $x\in\ker(M)$;
(ii) $N^\top x=0$, which implies $x\in\ker(N^\top)$.

Thus, one has that (21) holds if and only if $x\in\ker(M)\cup\ker(N^\top)$. Therefore, (21) holds $\forall x\in\mathbb{R}^n$ if and only if

$$\ker(M)\cup\ker(N^\top)=\mathbb{R}^n. \tag{22}$$

Since $\mathbb{R}^n$ is a subspace, and the union of two subspaces is a subspace if and only if one is a subset of the other [17], it results that (22) holds if and only if one of the following two statement hold:

(I) $\ker(M)\subseteq\ker(N^\top)=\mathbb{R}^n$;
(II) $\ker(N^\top)\subseteq\ker(M)=\mathbb{R}^n$.

Statement (I) cannot hold due to the fact that N is not the zero matrix. Therefore, if (21) holds for all $x\in\mathbb{R}^n$, then (II) holds, *i.e.*, M is the zero matrix. $\square$

By relying on Lemmas 1 and 2, we can prove Theorem 1.

Sufficiency: In order to show sufficiency of the stated conditions, define the quadratic functions $V(e)=e^\top Pe$, substitute the *information-constrained* control laws (6) in (3), and consider the Hamilton-Jacobi-Bellman equation [25]

$$0=\min_K\{e^\top(2P(A+KC)+Q+C^\top K^\top RKC)e\}, \tag{23}$$

which should hold for any $e\in\mathbb{R}^n$. By [25], there exists a solution to the optimal observation problem (3), (4), provided there exist P that solve such equation. Therefore, define $M(K)=2PKC+C^\top K^\top RKC$, and note that

$$\tfrac{\partial}{\partial K}(2P(A+KC)+Q+C^\top K^\top RKC)=\tfrac{\partial}{\partial K}M(K).$$

By borrowing the tools discussed in [22], one obtains that

$$\begin{aligned}
\tfrac{\partial(e^\top M(K)e)}{\partial K} &= 2(I_n\otimes e^\top P)\overline{U}_{n\times p}(I_p\otimes Ce)\\
&\quad+(I_n\otimes e^\top C^\top)U_{n\times p}(I_p\otimes RKCe)\\
&\quad+(I_n\otimes e^\top C^\top K^\top R)\overline{U}_{n\times p}(I_p\otimes Ce).
\end{aligned}$$

The only terms dependent on K in the expression above are

$$\begin{aligned}
N(K) &= (I_n\otimes e^\top C^\top)U_{n\times p}(I_p\otimes RKCe)\\
&\qquad+(I_n\otimes C^\top K^\top R)\overline{U}_{n\times p}(I_p\otimes C).
\end{aligned}$$

Furthermore, by Lemma 1, it follows that

$$\begin{aligned}
&(I_n\otimes e^\top C^\top)U_{n\times p}(I_p\otimes RKCe)\\
&\quad=(e^\top C^\top\otimes I_n)(I_p\otimes RKCe)=RKCee^\top C^\top\\
&\quad=(I_n\otimes e^\top C^\top K^\top R)\overline{U}_{n\times p}(I_p\otimes Ce).
\end{aligned}$$

Therefore, $N(K)$ can be rewritten as $N(K)=2(I_n\otimes e^\top C^\top)U_{n\times p}(I_p\otimes RKCe)$. Consider now the Hessian matrix H of the (scalar) function $e^\top M(K)e$ with respect to $\mathrm{vec}(K)$,

$$\begin{aligned}
H &:= \tfrac{\partial}{\partial(\mathrm{vec}(K))^\top}\left(\mathrm{vec}\left(\tfrac{\partial}{\partial K}(e^\top M(K)e)\right)\right)\\
&\qquad= \tfrac{\partial}{\partial(\mathrm{vec}(K))^\top}\left(\mathrm{vec}(N(K))\right).
\end{aligned}$$

By the reasoning given above, it results that $\mathrm{vec}(N(K)) = 2\,\mathrm{vec}((e^\top C^\top\otimes I_n)(I_p\otimes RKCe)) = 2\,\mathrm{vec}(RKCee^\top C^\top) = 2((Cee^\top C^\top)\otimes R)\,\mathrm{vec}(K)$. Hence, one has that $H = 2(Cee^\top C^\top)\otimes R$, which is positive semidefinite due to the positive semidefiniteness of $Cee^\top C^\top$ and R. Hence, the scalar function $e^\top M(K)e$ is convex with respect to $\mathrm{vec}(K)$. Therefore, by classical results about convex functions [26], the matrix $K^\star$ minimizes $e^\top M(K)e$ for all $e \in \mathbb{R}^n$ if and only if $\frac{\partial}{\partial K}M(K^\star) = 0$. Thus, by Lemma 1, $K^\star$ must satisfy $RK^\star Cee^\top C^\top = -Pee^\top C^\top$, for all $e \in \mathbb{R}^n$, *i.e.*,

$$(RK^\star C + P)ee^\top C^\top = 0. \tag{24}$$

Thus, by (24) and Lemma 2, one has that $K^\star$ must be such that $RK^\star C = -P$. By [22], (24) can be written as

$$\mathrm{vec}(RK^\star C) = (C^\top \otimes R)\,\mathrm{vec}(K^\star) = -\,\mathrm{vec}(P), \tag{25}$$

i.e., $\mathrm{vec}(P)$ must be in the image of $C^\top \otimes R$, *i.e.*, (7b) must hold. Thus, substituting $RK^\star C$ with $-P$ into (23), we obtain $e^\top(PA + A^\top P + Q - PR^\dagger P)e = 0$ for all $e \in \mathbb{R}^n$, *i.e.*, P must satisfy (7a). Therefore, if there exists P that solves (7a) and such that (7b) holds, then, letting $K^\star$ be a solution to (25), one has that $v^\star = K^\star e$ is an admissible optimal correction input for (3), (4).

Necessity: In order to prove necessity, assume that there is an admissible optimal correction input $v^\star = K^\star e$, but (7a) does not hold or $\mathrm{vec}(P) \notin \mathrm{im}\left(C^\top \otimes R\right)$. Define the function $V(e) = J^\star(e)$, which is analytic by (12) and [19] and hence must satisfy [25], for all $e \in \mathbb{R}^n$,

$$0 = \tfrac{\partial V}{\partial e}(A + K^\star C)e + e^\top Qe + e^\top C^\top K^{\star\top} RK^\star Ce.$$

By considering the Taylor series expansion of V about the origin $e = 0$, one has that V can be expressed as $V = \sum_{\ell \geqslant 1} p_\ell$, where p_ℓ is a homogeneous polynomial in e of degree ℓ. Therefore, since $\frac{\partial}{\partial e}V = \sum_{\ell \geqslant 1}\frac{\partial}{\partial e}p_\ell$ and p_ℓ is homogeneous of degree ℓ with respect to the standard dilation [27], one has that $\frac{\partial p_\ell}{\partial e}(A + K^\star C)e$ is still homogeneous of degree ℓ. Thus, letting $A^\star = A + K^\star C$, the expression given in (26) can be equivalently rewritten as

$$\begin{aligned} &\tfrac{\partial p_\ell}{\partial e}A^\star e = 0, && \ell \geqslant 1, \ell \neq 2,\\ &\tfrac{\partial p_2}{\partial e}A^\star e = -e^\top Qe - e^\top C^\top (K^\star)^\top RK^\star Ce, \end{aligned}$$

Thus, by rewriting p_2 as $p_2 = e^\top Pe$, for some symmetric $P \in \mathbb{R}^{n\times n}$, one has that (23) must hold, thus leading to a contradiction by the reasoning given to prove sufficiency.

B. Proof of Theorem 2

Sufficiency: It suffices to show that if there exist symmetric and positive definite matrices Q and R such that (16) holds then the hypotheses of Theorem 1 are met. Hence, assuming that (16) holds, let $P = R\Pi^\diamond S$, which is symmetric and positive semidefinite by (16d) and (16c). Furthermore, by (16e), such an equation solves (7a). Thus, it remains to prove that (7b) holds. Note that, by construction, we have

$$\begin{aligned} \mathrm{vec}(P) &= \mathrm{vec}(R\Pi^\diamond S) = \mathrm{vec}(R\Pi^\diamond C^\top W^{-1} C)\\ &= (C^\top \otimes R)\,\mathrm{vec}(\Pi^\diamond C^\top W^{-1}), \end{aligned}$$

and hence also (7b) holds, thus concluding the proof.

Necessity: To prove necessity, assume that the steady-state behavior of the Kalman-Bucy filter is optimal for some cost functional of the form (4), with Q and R substituted by $Q^\diamond$ and $R^\diamond$, respectively, which are not both zero, but that there does not exist a nontrivial solution to (16). By Theorem 1, this implies that there exists $P^\diamond = P^{\diamond\top}$, $P^\diamond \succeq 0$, that solves (7) and such that $P^\diamond = R^\diamond \Pi^\diamond C^\top W^{-1} C$. This implies that $Q^\diamond$ and $R^\diamond$ solve (16), thus leading to a contradiction.

References

[1] D. G. Luenberger, "Observing the state of a linear system," *IEEE Trans. Mil. Electron.*, vol. 8, no. 2, pp. 74–80, 1964.

[2] D. Luenberger, "Observers for multivariable systems," *IEEE Trans. Autom. Control*, vol. 11, no. 2, pp. 190–197, 1966.

[3] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, vol. 83, no. 1, pp. 95–108, 1961.

[4] J. O'Reilly, *Observers for linear systems*. Academic Press, 1983.

[5] C. Possieri and A. R. Teel, "Structural properties of a class of linear hybrid systems and output feedback stabilization," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2704 – 2719, 2017.

[6] E. F. Costa and A. Astolfi, "On the stability of the recursive Kalman filter for linear time-invariant systems," in *Am. Control Conf.*, pp. 1286–1291, 2008.

[7] V. I. Utkin, *Sliding modes in optimization and control problems*. Springer Verlag, New York, 1992.

[8] A. Popescu, G. Besançon, and A. Voda, "A new robust observer approach for unknown input and state estimation," in *Eur. Control Conf.*, pp. 1607–1612, 2018.

[9] H. Gao, X. Meng, and T. Chen, "A new design of robust H_2 filters for uncertain systems," *Syst. Control Lett.*, vol. 57, no. 7, pp. 585 – 593, 2008.

[10] L. Wang and A. Morse, "A distributed observer for a time-invariant linear system," *IEEE Trans. Autom. Control*, 2017.

[11] H. Ríos and A. R. Teel, "A hybrid fixed-time observer for state estimation of linear systems," *Automatica*, vol. 87, pp. 103–112, 2018.

[12] J. P. Hespanha, *Linear systems theory*. Princeton Univ. Press, 2018.

[13] H. Kwakernaak and R. Sivan, *Linear optimal control systems*. Wiley-Interscience New York, 1972.

[14] J. C. Willems, "Deterministic least squares filtering," *J. Econom.*, vol. 118, no. 1-2, pp. 341–373, 2004.

[15] A. Ferrante and L. Ntogramatzidis, "Continuous-time singular linear–quadratic control: Necessary and sufficient conditions for the existence of regular solutions," *Syst. Control Lett.*, vol. 93, pp. 30–34, 2016.

[16] J. C. Willems, A. Kitapci, and L. M. Silverman, "Singular optimal control: a geometric approach," *SIAM J. Control Optim.*, vol. 24, no. 2, pp. 323–337, 1986.

[17] C. D. Meyer, *Matrix analysis and applied linear algebra*. SIAM, 2000.

[18] P. Dorato, V. Cerone, and C. Abdallah, *Linear-quadratic control: an introduction*. Simon & Schuster, 1994.

[19] P. Dorato, L. Menini, and C. Treml, "Robust multi-objective feedback design with linear guaranteed-cost bounds," *Automatica*, vol. 34, no. 10, pp. 1239 – 1243, 1998.

[20] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge Univ. press, 1990.

[21] M. J. Grimble and M. A. Johnson, *Optimal control and stochastic estimation: theory and applications*. John Wiley & Sons, Inc., 1986.

[22] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, 1978.

[23] X. Magnus and H. Neudecker, *Matrix differential calculus*. John Wiley & Sons, 1988.

[24] D. Turkington, "Generalised vec operators and the seemingly unrelated regression equations model with vector correlated disturbances," *J. Econometrics*, vol. 99, no. 2, pp. 225–253, 2000.

[25] D. P. Bertsekas, *Dynamic programming and optimal control*, vol. I, II. Athena scientific, 2012.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[27] L. Menini and A. Tornambe, *Symmetries and Semi-invariants in the Analysis of Nonlinear Systems*. Springer, 2011.