

Linear-quadratic Gaussian Games with Asymmetric Information: Belief Corrections Using the Opponents Actions^{*}

Ben Hambly[†]

Renyuan Xu[‡]

Huining Yang[§]

August 4, 2023

Abstract

We consider two-player non-zero-sum linear-quadratic Gaussian games in which both players aim to minimize a quadratic cost function while controlling a linear and stochastic state process using linear policies. The system is partially observable with asymmetric information available to the players. In particular, each player has a private and noisy measurement of the state process but can see the history of their opponent's actions. The challenge of this asymmetry is that it introduces correlations into the players' belief processes for the state and leads to circularity in their beliefs about their opponents beliefs. We show that by leveraging the information available through their opponent's actions, both players can enhance their state estimates and improve their overall outcomes. In addition, we provide a closed-form solution for the Bayesian updating rule of their belief process. We show that there is a Nash equilibrium which is linear in the estimation of the state and with a value function incorporating terms that arise due to errors in the state estimation. We illustrate the results through an application to bargaining which demonstrates the value of these information corrections.

1 Introduction

In the realm of game theory, two-player bargaining games have been extensively studied due to their practical implications in various domains [16, 4, 21, 9]. The types of game of interest here typically occur in negotiations between large organizations. One example is the negotiation of extraction rights between a host country and a foreign mining company. The host country seeks to maximize its economic benefits while ensuring sustainable resource management, while the foreign company aims to secure profitable extraction agreements. Another relevant scenario is the negotiation of rebates on pharmaceutical purchases between a manufacturer and a retail chain. The manufacturer desires to maximize its sales volume and market share, while the retail chain seeks competitive pricing to attract customers. To capture these situations we consider negotiations over a good whose value changes through time and assume that there are a fixed number of rounds of negotiation, where at each round both players present their bids simultaneously. In practice, both players possess limited and potentially asymmetric information about the good under negotiation. This provides compelling motivation for the mathematical modeling and analysis of two-player stochastic games where there is asymmetric information, which can shed light on the negotiation dynamics, identify the key parameters affecting the bargaining outcome, and provide insights into the search for information.

^{*}We thank Sam Cohen and Yanwei Jia for helpful discussions and comments on the paper.

[†]Mathematical Institute, University of Oxford. **Email:** hambly@maths.ox.ac.uk

[‡]Epstein Department of Industrial and Systems Engineering, University of Southern California. **Email:** renyuanx@usc.edu

[§]Department of Operations Research and Financial Engineering, Princeton University. **Email:** hy5564@princeton.edu

In the game setting where the underlying state dynamics are only partially observed, the players’ observation histories expand over time, leading to strategies with *expanding domains* due to the dynamic and sequential nature of such games. To address this issue, a commonly employed technique is to summarize the time-expanding histories into *sufficient statistics*. This enables the application of the Dynamic Programming Principle (DPP), allowing such sequential decision-making problems to be solved through nested sub-optimization problems using the Bellman equation. The sufficient statistics may vary from player to player, depending on the observations available to each player. In the case of symmetric information in extensive form games, the concept of a Markov perfect equilibrium has been introduced [10], in which the players’ strategies are determined solely by past events that are relevant to their payoffs, rather than the entire history. See also [3, 15]. However, for games with asymmetric information, identifying the appropriate sufficient statistics poses a significant challenge.

A quantity commonly used as the sufficient statistic is a *posterior belief* of the state dynamics, constructed using the available observations and Bayes formula. The main difficulty in this context is the emergence of *private beliefs*, the fact that different agents in the system may have different (private) observations about the same unknown quantity, which introduces *dependence* among the agents’ beliefs. One way to avoid this problem is to consider models in which private beliefs either do not exist (such as considering symmetric information games, or asymmetric but independent observations [18, 19]), or, if they do exist, they are not taken into account in the agents’ strategies (see for example the concept of “public perfect equilibrium” [1]). Another closely related line of work considers common-information based Markov perfect equilibria, which breaks the history into the common and private parts. This idea was first introduced in [11] for finite games and then generalized in [7] to linear-quadratic games. See [12] for an extension to a more general setting and [6] for an application to cyber-physical systems. One key assumption made in this approach is that players’ posterior beliefs about the system state, conditioned on their common information, are independent of the strategies used by the players in the past. This decouples the sequential rationality and belief consistency, resulting in a simplification in calculating the equilibria, and obviating the need to define (possibly correlated) private beliefs.

To better understand the challenge posed by having private beliefs, let us consider the following simplified scenario. We have two players, P and E , who collect private information about an unknown variable Θ at each time step. Player P acts based on her own private belief about Θ , and expects that Player E will do the same. Although both players can observe each other’s actions, they cannot observe each other’s private beliefs. This means that Player P must form a belief about Player E ’s private belief in order to interpret the actions from player E , and take this into account when making her own decisions. However, this creates a chain of “belief about belief” that must be taken into consideration, which extends as long as each player’s beliefs remain private. Due to the symmetry of the information structure, Player E must do the same. Thus, Player P needs to form beliefs about beliefs about beliefs of Player E , creating an increasingly complex hierarchy of beliefs. This chain only stops when a belief in one step becomes a public function of the beliefs in the previous steps.

Indeed, stochastic games with private beliefs have been identified as an open problem in the past decade [13, 14]. There have been a few attempts to address this challenging issue. In [8] a model with an unknown (but static) state Θ of the world is considered, where each player has a private noisy observation of Θ at timestamp t . The private observations are independent given Θ . The authors specialize this setting to the case of a linear-quadratic Gaussian non-zero-sum game where Θ is a Gaussian variable and players’ observations are generated through a linear Gaussian model from Θ . The main contribution of the paper is to show that, due to the conditional independence of the private signals given Θ , the private belief chain stops at the second step and players’ beliefs about others’ beliefs are public functions of their own beliefs (the first step beliefs). In addition, the authors show that the perfect Bayesian equilibrium (PBE) for their model is a linear function of a players’ private Kalman filter estimator. For a more general setting when the (partially) unknown quantity is the underlying stochastic process itself, [13] considers a zero-sum two-player game formulation in which

each agent observes a private linear signal of the underlying process with non-degenerate Gaussian noise. The private signals at each time step are independent conditioned on the true value of the state at the same time. Similarly, the author shows that the private belief chain stops at the second step. However, the sufficient statistics developed in [13] are not completely correct, impeding the application of the DPP and making it impossible to derive the Nash equilibrium solutions. We will give a detailed technical discussion in Section 2.2.

Our Contributions. Motivated by bargaining games, we consider a two-player non-zero-sum game with linear dynamics and quadratic cost functions. Both players cannot directly observe the dynamics but instead rely on private signal processes that are linear in the state process with some additive Gaussian noise. In addition, both players adopt linear policies to control the partially observable dynamics and each player can also observe the past actions taken by the other player.

The novelty of our approach is that we show how to use the opponent's actions as additional information to correct the state estimate of the previous time step via a modified Kalman filtering procedure. For the partially observable setting, we formally derive the updating rule for the belief process and provide an explicit formula for the projection of the unknown state process onto the filtration generated by the information flow that is available to each player in Theorem 2.4. In addition, we prove a conditional version of the DPP that works in the game setting in Theorem 3.1. With the sufficient statistics and DPP in hand, we are able to derive a Nash equilibrium solution for the two-player game in Theorem 3.4. Finally, in Section 4, we extend the above-mentioned results to a more general setting where part of the state is fully observable and part of the state is partially observable through the signal process. In the mixed partially and fully observable setting, we establish parallel findings—specifically, the updating rule for the sufficient statistics in Theorem 4.2 and the Nash equilibrium in Theorem 4.3. To the best of our knowledge, this is the first work that rigorously characterizes the equilibrium solution under private belief when the underlying state dynamics are partially observable. To demonstrate the performance of our framework, we conclude the paper in Section 5 by discussing a bargaining game example and give a numerical illustration of the effects of using information corrections.

2 Problem Set-up

We consider a general setting for games with two players, P and E , under partial observations and asymmetric information. The joint dynamics of the state process $x_t \in \mathbb{R}^n$ takes a linear form ($0 \leq t \leq T-1$):

$$x_{t+1} = A_t x_t + B_t^P u_t^P + B_t^E u_t^E + \Gamma_t w_t, \quad (2.1)$$

with initial value $x_0 = x$, and the controls of P and E are $u_t^P \in \mathbb{R}^m$ and $u_t^E \in \mathbb{R}^k$, respectively. Here, for each t , the process noise $w_t \in \mathbb{R}^d$ is an i.i.d. sample from $\mathcal{N}(0, W)$ with $W \in \mathbb{R}^{d \times d}$ and we have the model parameters $A_t \in \mathbb{R}^{n \times n}$, $B_t^P \in \mathbb{R}^{n \times m}$, $B_t^E \in \mathbb{R}^{n \times k}$, and $\Gamma_t \in \mathbb{R}^{n \times d}$.

Information Structure. At the time $t = 0$, player P is not able to observe x_0 but instead believes that the initial state is drawn from a Gaussian distribution. Namely, from the view point of player P ,

$$x_0 \sim \mathcal{N}(\hat{x}_0^P, W_0^P), \quad (2.2)$$

where \hat{x}_0^P is their own initial constant and W_0^P is a known constant covariance matrix. After that, player P observes the following noisy state signal (or measurement) $z_t^P \in \mathbb{R}^p$:

$$z_{t+1}^P = H_{t+1}^P x_{t+1} + w_{t+1}^P, \quad w_{t+1}^P \sim \mathcal{N}(0, G^P), \quad t = 0, 1, \dots, T-1, \quad (2.3)$$

with $\{w_t^P\}_{t=0}^{T-1}$ a sequence of i.i.d. random variables. Here $G^P \in \mathbb{R}^{p \times p}$ and $H_{t+1}^P \in \mathbb{R}^{p \times n}$. Similarly, player E believes that the initial state is drawn from a Gaussian distribution:

$$x_0 \sim \mathcal{N}(\hat{x}_0^E, W_0^E), \quad (2.4)$$

with their own initial constant \hat{x}_0^E and known constant W_0^E . Thereafter player E observes the following noisy state signal $z_t^E \in \mathbb{R}^q$:

$$z_{t+1}^E = H_{t+1}^E x_{t+1} + w_{t+1}^E, \quad w_{t+1}^E \sim \mathcal{N}(0, G^E), \quad t = 0, 1, \dots, T-1. \quad (2.5)$$

with $\{w_t^E\}_{t=0}^{T-1}$ a sequence of i.i.d. random variables. For simplicity we assume that $\{w_t^E\}_{t=0}^{T-1}$ are independent from $\{w_t^P\}_{t=0}^{T-1}$. In addition, $G^E \in \mathbb{R}^{q \times q}$ and $H_{t+1}^E \in \mathbb{R}^{q \times n}$.

We follow [13] to define games with perfect, imperfect, and partial observations:

- (a) If the observation matrices H_t^P and H_t^E are both the identity matrix and there is no measurement noise, that is, the covariance matrices $G^P = 0$ and $G^E = 0$, we have a game with *full (or perfect) observation*. In other words, the players' observation is $z_t^E = z_t^P = x_t$.
- (b) If the observation matrices H_t^P and H_t^E are the identity matrix and there is measurement noise, we have a game with *imperfect observation*. Namely, the players' observations are of the form

$$z_t^P = x_t + w_t^P, \quad z_t^E = x_t + w_t^E.$$

- (c) If the observation matrices are not the identity and there is measurement noise, as in

$$z_t^P = H_t^P x_t + w_t^P, \quad z_t^E = H_t^E x_t + w_t^E,$$

we have a game with *partial observation*.

Both players make their decisions based on the public and private information available to them. We write $\mathcal{Z}_t^P = \{z_s^P\}_{s=1}^t$ and $\mathcal{Z}_t^E = \{z_s^E\}_{s=1}^t$ for the private signals players P and E receive up to time t ($1 \leq t \leq T$), respectively. Let $\mathcal{U}_t^P = \{u_s^P\}_{s=1}^t$ and $\mathcal{U}_t^E = \{u_s^E\}_{s=1}^t$ denote the control history of player P and player E up to time t , respectively.

We assume \mathcal{H}_t^P is the information (or history) available to player P and \mathcal{H}_t^E is the information available to player E for them to make decisions at time t , where \mathcal{H}_t^P and \mathcal{H}_t^E follow:

$$\mathcal{H}_t^P = \{\hat{x}_0^P, W_0^P, W_0^E\} \cup \mathcal{Z}_t^P \cup \mathcal{U}_{t-1}^P \cup \mathcal{U}_{t-1}^E, \quad \mathcal{H}_t^E = \{\hat{x}_0^E, W_0^P, W_0^E\} \cup \mathcal{Z}_t^E \cup \mathcal{U}_{t-1}^P \cup \mathcal{U}_{t-1}^E. \quad (2.6)$$

Note that the covariance matrices $\{W_0^P, W_0^E\}$ are known to both players. In the posterior update of a Gaussian distribution, sufficient statistics involve both the mean and covariance matrix. Knowing $\{W_0^P, W_0^E\}$ is essential for both players to update their posterior covariance matrices. In addition, we highlight that both players know all the model parameters.

Cost Function. We consider a non-zero sum game between player P and player E where they strive to minimize their own cost functions. Player i 's ($i = P, E$) cost function is given by

$$\min_{\{u_t^i\}_{t=0}^{T-1}} J^i(\hat{x}_0^i) := \min_{\{u_t^i\}_{t=0}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^i x_T + \sum_{t=0}^{T-1} \left(x_t^\top Q_t^i x_t + (u_t^i)^\top R_t^i u_t^i \right) \middle| \mathcal{H}_0^i \right], \quad (2.7)$$

with cost parameters $Q_t^P, Q_t^E \in \mathbb{R}^{n \times n}$, $R_t^P \in \mathbb{R}^{m \times m}$, and $R_t^E \in \mathbb{R}^{k \times k}$.

For the well-definedness of the game, we summarize the assumptions on the model parameters, initial state, and noise.

Assumption 2.1 (Parameters, Initial State, and Noise). *For $i = P, E$,*

1. $\{w_t\}_{t=0}^{T-1}$ and $\{w_t^i\}_{t=1}^{T-1}$ are zero-mean, i.i.d. Gaussian random variables that are independent from x_0 and each other and such that $\mathbb{E}[w_t w_t^\top] = W$ is positive definite and $\mathbb{E}[w_t^i (w_t^i)^\top] = G^i$ is positive definite;
2. Both matrices $H_{t+1}^P \in \mathbb{R}^{p \times n}$ and $H_{t+1}^E \in \mathbb{R}^{q \times n}$ have rank n for $t = 0, \dots, T-1$.
3. The matrices $\Gamma_t W \Gamma_t^\top$ are non-singular for $t = 1, \dots, T$;
4. The cost matrices Q_t^i , for $t = 0, 1, \dots, T$ are positive semi-definite, and R_t^i for $t = 0, 1, \dots, T-1$ are positive definite.

Under the assumptions we make on G^P and G^E , we exclude the perfect information case as we require G^P and G^E to be positive definite. This is a challenging scenario to study as the agent cannot get precise information for any coordinate of the state process. On the other hand, since $H_{t+1}^P \in \mathbb{R}^{p \times n}$ and $H_{t+1}^E \in \mathbb{R}^{q \times n}$ have rank n , the signal process will reveal aggregated information from all of the coordinates. Hence the agent is still capable of gradually learning each coordinate of the state process.

Remark 2.2. We focus on the case where all states are partially observable, as given in Assumption 2.1, for the rest of Section 2 and for all of Section 3. We generalize it to a mixed fully and partially observable setting in Section 4.

2.1 Sufficient Statistics: Decentralized Kalman Filtering with Information Correction

To better demonstrate the sufficient statistics of Kalman filtering in the game setting, we start with a brief discussion of some existing results in the single-agent setting.

2.1.1 Preliminary: Single-agent setting

Suppose there is a single player P who controls the state dynamics (2.1):

$$x_{t+1} = A_t x_t + B_t^P u_t^P + \Gamma_t w_t, \quad (2.8)$$

where $x_0 = x$ is the initial position and $u_t^P \in \mathbb{R}^m$ are the controls from player P .

Information Structure. In this single-player case player P believes that the initial state is drawn from a Gaussian distribution at the time $t = 0$:

$$x_0 \sim \mathcal{N}(\hat{x}_0^P, W_0^P), \quad (2.9)$$

and thereafter player P observes the following noisy state signal $z_t \in \mathbb{R}^p$:

$$z_{t+1}^P = H_{t+1}^P x_{t+1} + w_{t+1}^P, \quad w_{t+1}^P \sim \mathcal{N}(0, G^P), \quad t = 0, 1, \dots, T-1, \quad (2.10)$$

with $\{w_t^P\}_{t=0}^{T-1}$ a sequence of i.i.d. random variables. Here $G^P \in \mathbb{R}^{p \times p}$ and $H_{t+1}^P \in \mathbb{R}^{p \times n}$.

Assume the information available to player P to make a decision at time t follows:

$$\mathcal{H}_t^P = \{\hat{x}_0^P, W_0^P\} \cup \mathcal{Z}_t^P \cup \mathcal{U}_{t-1}^P, \quad (2.11)$$

then we have the following result characterizing player P 's belief in the state.

Theorem 2.3. [17, (5.3-39)-(5.3-42)] *The sufficient statistic for player P at decision time $t = 0$ is (\hat{x}_0, W_0^P) . Namely player P believes that $x_0 \sim \mathcal{N}(\hat{x}_0, W_0^P)$. For time $1 \leq t \leq T - 1$, the distribution of the physical state x_t calculated by player P , by conditioning on the private information available to him at time t , is given by*

$$x_t \sim \mathcal{N}(\hat{x}_t^P, \hat{\Sigma}_t^P), \quad (2.12)$$

where

$$(\hat{x}_t^P)^- = A_{t-1}\hat{x}_{t-1}^P + B_{t-1}^P u_{t-1}^P, \quad (2.13a)$$

$$(\hat{\Sigma}_t^P)^- = A_{t-1}\hat{\Sigma}_{t-1}^P A_{t-1}^\top + \Gamma_{t-1} W \Gamma_{t-1}^\top, \quad (2.13b)$$

$$K_t^P = (\hat{\Sigma}_t^P)^- (H_t^P)^\top \left[H_t^P (\hat{\Sigma}_t^P)^- (H_t^P)^\top + G^P \right]^{-1}, \quad (2.13c)$$

$$\hat{x}_t^P = (\hat{x}_t^P)^- + K_t^P \left[z_t^P - H_t^P (\hat{x}_t^P)^- \right], \quad (2.13d)$$

$$\hat{\Sigma}_t^P = (I - K_t^P H_t^P) (\hat{\Sigma}_t^P)^-, \quad (2.13e)$$

with initial condition $\hat{x}_0^P = \hat{x}_0$ and $\hat{\Sigma}_0^P = W_0^P$.

Note that $(\hat{x}_t^P)^- = \mathbb{E} \left[x_t \middle| \mathcal{H}_{t-1} \right]$ is the state estimate obtained before the measurement update, namely, using information up to time $t - 1$. This term is often called the *pre-estimate* in the literature. With the new measurement information z_t^P at time t , the agent updates the state estimate to $\hat{x}_t^P = \mathbb{E} \left[x_t \middle| \mathcal{H}_t^P \right]$, which is a linear combination of $(\hat{x}_t^P)^-$ and z_t^P . This term is often called the *post-estimate* [17]. Here the second term $z_t^P - H_t^P (\hat{x}_t^P)^-$ on the RHS of (2.13d) is independent of the first term $(\hat{x}_t^P)^-$. The only decision variable K_t^P in the coefficients is chosen so that the conditional mean-square error is minimized. Namely,

$$K_t^P = \arg \min \mathbb{E} \left[\|x_t - \hat{x}_t^P\|^2 \middle| \mathcal{H}_t^P \right]. \quad (2.14)$$

In terms of quantifying the uncertainty in the state estimate, we have $(\hat{\Sigma}_t^P)^- = \mathbb{E} \left[(x_t - (\hat{x}_t^P)^-)^\top (x_t - (\hat{x}_t^P)^-) \right]$ and $\hat{\Sigma}_t^P = \mathbb{E} \left[(x_t - \hat{x}_t^P)^\top (x_t - \hat{x}_t^P) \right]$ representing the covariance before and after measurement updates, respectively.

2.1.2 Two-player setting

There are a few challenges for the belief updates in the game setting. In this paper we focus on the case where both players P and E adopt linear feedback policies:

$$u_t^P := F_t^P \mathbb{E}[x_t | \mathcal{H}_t^P], \text{ and } u_t^E := F_t^E \mathbb{E}[x_t | \mathcal{H}_t^E], \quad (2.15)$$

with some *policy matrices* $F_t^P \in \mathbb{R}^{m \times n}$ and $F_t^E \in \mathbb{R}^{k \times n}$.

From player P 's perspective, the new information collected at time t is z_t^P and u_{t-1}^E . Intuitively, player P should be able to use u_{t-1}^E as some additional information to improve their estimate of x_{t-1} . Note that player P is aware that player E adopts a linear feedback policy $u_{t-1}^E = F_{t-1}^E \mathbb{E}[x_{t-1} | \mathcal{H}_{t-1}^E]$, for which the opponent's policy matrix F_{t-1}^E (a function of model parameters) is known but the state estimate $\mathbb{E}[x_{t-1} | \mathcal{H}_{t-1}^E]$ is unknown to player P . In order to utilize the information contained in u_{t-1}^E , player P also needs to infer the distribution of $\mathbb{E}[x_{t-1} | \mathcal{H}_{t-1}^E]$ using \mathcal{H}_{t-1}^P . We will show later on in Section 2.2 that this idea of *using the opponent's actions to make an information correction* is not only

a possible approach to improve the estimation precision but also a necessary step to guarantee that the conditional expectations based on the information filtrations satisfy the tower property and hence that the DPP holds.

Let us start at decision time $t = 0$. Player P reasons as follows. He models his initial belief \hat{x}_0^P of the initial state x_0 as

$$\hat{x}_0^P = x_0 + e_0^P, \quad (2.16)$$

where x_0 is the true physical state and e_0^P is player P 's estimation error, whose distribution, in view of (2.2), is $\mathcal{N}(0, W_0^P)$. In the same way, player E 's belief \hat{x}_0^E of the initial state x_0 follows

$$\hat{x}_0^E = x_0 + e_0^E, \quad (2.17)$$

where, as before, x_0 is the true physical state and e_0^E is player E 's estimation error, whose distribution, in view of (2.4), is $\mathcal{N}(0, W_0^E)$. The Gaussian random variables e_0^E and e_0^P are independent by our assumptions.

From player P 's perspective, \hat{x}_0^P is known but \hat{x}_0^E is a random variable. Subtracting (2.16) from (2.17), at time $t = 0$ player P concludes that as far as he is concerned, player E 's estimate, upon which he will decide his optimal control, is the random variable

$$\hat{x}_0^E = \hat{x}_0^P + e_0^E - e_0^P. \quad (2.18)$$

As far as P is concerned, E 's estimate of the initial state x_0 is a Gaussian random variable

$$\hat{x}_0^E \sim \mathcal{N}(\hat{x}_0^P, W_0^P + W_0^E). \quad (2.19)$$

Thus, at time $t = 0$ player P has used his private information \hat{x}_0^P and the public information (W_0^P, W_0^E) to calculate the distribution of the sufficient statistic \hat{x}_0^E of player E . Similarly, as far as player E is concerned, at time $t = 0$ the distribution of the initial state estimate \hat{x}_0^P of player P follows $\mathcal{N}(\hat{x}_0^E, W_0^P + W_0^E)$.

In addition to (2.15), we further restrict the admissible set of policy matrices to the following:

$$\mathcal{A}^P := \{F^P \in \mathbb{R}^{m \times n} | F^P \text{ has rank } \min(m, n)\}, \quad \mathcal{A}^E := \{F^E \in \mathbb{R}^{k \times n} | F^E \text{ has rank } \min(k, n)\}.$$

For time $t \geq 1$, we have the following result.

Theorem 2.4 (Sufficient Statistics in Two-player Games). *Assume the sufficient statistic of player i ($i = P, E$) at decision time $t = 0$ is $x_0 \sim \mathcal{N}(\hat{x}_0^i, W_0^i)$. In addition, assume both players are applying linear strategies. Namely, $u_t^P = F_t^P \mathbb{E}[x_t | \mathcal{H}_t^P]$ and $u_t^E = F_t^E \mathbb{E}[x_t | \mathcal{H}_t^E]$ for some matrices $F_t^P \in \mathcal{A}^P$ and $F_t^E \in \mathcal{A}^E$. Then, for time $1 \leq t \leq T - 1$, the distribution of the physical state x_t calculated by player i conditioning on the private information available to him at time t follows*

$$x_t \sim \mathcal{N}(\hat{x}_t^i, \hat{\Sigma}_t^i), \quad (2.20)$$

where, for $j \neq i$,

$$J_{t-1}^i = \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right) \hat{\Sigma}_{t-1}^{(i,j)} (Y_{t-1}^j)^\top \left(Y_{t-1}^j \hat{\Sigma}_{t-1}^{(i,j)} \hat{\Sigma}_{t-1}^{(i,j)} (Y_{t-1}^j)^\top \right)^{-1}, \quad (2.21a)$$

$$(\hat{x}_{t-1}^i)^+ = \hat{x}_{t-1}^i + J_{t-1}^i (y_{t-1}^j - Y_{t-1}^j \hat{x}_{t-1}^i), \quad (2.21b)$$

$$(\hat{\Sigma}_{t-1}^i)^+ = \hat{\Sigma}_{t-1}^i - \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right) (\hat{\Sigma}_{t-1}^{(i,j)})^{-1} \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right)^\top, \quad (2.21c)$$

$$(\hat{x}_t^i)^- = A_{t-1} (\hat{x}_{t-1}^i)^+ + B_{t-1}^P u_{t-1}^P + B_{t-1}^E u_{t-1}^E, \quad (2.21d)$$

$$(\hat{\Sigma}_t^i)^- = A_{t-1} (\hat{\Sigma}_{t-1}^i)^+ A_{t-1}^\top + \Gamma_{t-1} W \Gamma_{t-1}^\top, \quad (2.21e)$$

$$K_t^i = (\hat{\Sigma}_t^i)^- (H_t^i)^\top \left[H_t^i (\hat{\Sigma}_t^i)^- (H_t^i)^\top + G^i \right]^{-1}, \quad (2.21f)$$

$$\hat{x}_t^i = (\hat{x}_t^i)^- + K_t^i \left[z_t^i - H_t^i (\hat{x}_t^i)^- \right], \quad (2.21g)$$

$$\hat{\Sigma}_t^i = (I - K_t^i H_t^i) (\hat{\Sigma}_t^i)^-, \quad (2.21h)$$

$$\tilde{\Sigma}_t^{(i,j)} = (I - K_t^i H_t^i) \left(A_{t-1} \Delta_{t-1}^{(i,j)} A_{t-1}^\top + \Gamma_{t-1} W \Gamma_{t-1}^\top \right) (I - K_t^j H_t^j)^\top \quad (2.21i)$$

$$\Delta_{t-1}^{(i,j)} = (\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)}) (\hat{\Sigma}_{t-1}^{(i,j)})^{-1} (\hat{\Sigma}_{t-1}^j - \tilde{\Sigma}_{t-1}^{(j,i)})^\top + \tilde{\Sigma}_{t-1}^{(i,j)} \quad (2.21j)$$

$$\hat{\Sigma}_t^{(i,j)} = \hat{\Sigma}_t^i + \hat{\Sigma}_t^j - \tilde{\Sigma}_t^{(i,j)} - \left(\tilde{\Sigma}_t^{(i,j)} \right)^\top, \quad (2.21k)$$

where $\hat{\Sigma}_{t-1}^{(i,j)}$ is positive definite. The values of $Y_t^P \in \mathbb{R}^{m \times n}$, $Y_t^E \in \mathbb{R}^{k \times n}$ and y_t^P, y_t^E depend on the ranks of F_t^P and F_t^E as follows:

(i) The pair

$$(Y_t^P, y_t^P) = \begin{cases} (F_t^P, u_t^P) & \text{if } F_t^P \text{ has rank } m < n, \\ (I_n, \hat{x}_t^P) & \text{if } F_t^P \text{ has rank } n \leq m. \end{cases}$$

(ii) The pair

$$(Y_t^E, y_t^E) = \begin{cases} (F_t^E, u_t^E) & \text{if } F_t^E \text{ has rank } k < n, \\ (I_n, \hat{x}_t^E) & \text{if } F_t^E \text{ has rank } n \leq k. \end{cases}$$

In addition, the initial conditions are $\hat{\Sigma}_0^i = W_0^i$, $\tilde{\Sigma}_0^{(i,j)} = 0$, and $\hat{\Sigma}_0^{(i,j)} = \hat{\Sigma}_0^i + \hat{\Sigma}_0^j$. Finally, in player i 's view, the posterior distribution for the state estimate \hat{x}_t^j of player j is

$$\hat{x}_t^j \sim \mathcal{N}(\hat{x}_t^i, \hat{\Sigma}_t^{(i,j)}). \quad (2.22)$$

- Remark 2.5.** 1. J_{t-1}^i in (2.21a) is the Kalman gain for player i when viewing player j 's action as the additional signal to improve the state estimation in the previous step $t-1$. We call $(\hat{x}_{t-1}^i)^+$ in (2.21b) the *improved-estimate* for x_{t-1} , with the corresponding estimation error $(\hat{\Sigma})_{t-1}^+$.
2. The post-estimates of the state and covariance after the measurement/signal update (2.21g)-(2.21h) take similar forms to the single-agent case (2.13d)-(2.13e). The differences occur in the input state and covariance estimates. In particular, the post-estimate for the single-agent setting uses the pre-estimate as the input whereas the post-estimate for the two-player setting uses the improved-estimate as the input.
3. Equation (2.22) in Theorem 2.4 shows that the chain of “belief about belief” stops at the second step, as the belief at the second step becomes a public function of the beliefs at the first step.
4. When $n > k, m$, players P and E will not be able to recover the opponent's state estimate via observing the action taken by the opponent. Instead, from player i 's viewpoint, the posterior distribution for the state estimate \hat{x}_t^j of player j follows a Gaussian distribution with mean \hat{x}_t^i and variance $\hat{\Sigma}_t^{(i,j)}$.
5. Consider the special case that

$$F_t^P \in \mathbb{R}^{m \times n} \text{ has rank } n, \ n \leq m, \text{ and } F_t^E \in \mathbb{R}^{k \times n} \text{ has rank } n, \ n \leq k. \quad (2.23)$$

In this case, player i can fully recover the state estimate from player j by observing her actions, as the RHS of the following equation is fully known to player i :

$$\mathbb{E}[x_t | \mathcal{H}_t^j] = ((F_t^j)^\top F_t^j)^{-1} (F_t^j)^\top u_t^j. \quad (2.24)$$

Observing that $J_t^i + J_t^j = I$, we have

$$(\hat{x}_t^i)^+ = (I - J_t^i) \hat{x}_t^i + J_t^i \hat{x}_t^j = J_t^j \hat{x}_t^i + (I - J_t^j) \hat{x}_t^j = (\hat{x}_t^j)^+.$$

This shows that Player P and Player E have the *same* improved estimate after observing each other's actions. In this case, information is fully shared between the players.

Proof. There are four possible combinations under the conditions (i)-(ii) stated in Theorem 2.4. Here we only show the proof for the following combination as the proof for each of the other combinations follows the same logic:

$$F_t^P \in \mathbb{R}^{m \times n} \text{ has rank } m, m < n; \text{ and } F_t^E \in \mathbb{R}^{k \times n} \text{ has rank } k, k < n. \quad (2.25)$$

In addition, under condition (2.25), we only prove the results for player P here as the results for player E follow in the same way.

We handle the new information u_{t-1}^E and z_t^P , in an incremental fashion. More precisely, we first adjust the estimate \hat{x}_{t-1}^P using u_{t-1}^E , denoted by $(\hat{x}_{t-1}^P)^+$, and then derive \hat{x}_t^P using z_t^P and $(\hat{x}_{t-1}^P)^+$.

After player P observes the action u_{t-1}^E from player E , player P updates:

$$(\hat{x}_{t-1}^P)^+ := \mathbb{E}[x_{t-1} | \mathcal{H}_{t-1}^P \cup \{u_{t-1}^E\}]. \quad (2.26)$$

Following the convention in filtering theory [17], we write:

$$(\hat{x}_{t-1}^P)^+ = \hat{x}_{t-1}^P + J_{t-1}^P (u_{t-1}^E - F_{t-1}^E \hat{x}_{t-1}^P), \quad (2.27)$$

where J_{t-1}^P is a matrix to be determined to minimize $\mathbb{E}[\|(x_{t-1}^P)^+ - x_{t-1}\|^2]$. To calculate J_{t-1}^P , we have

$$\begin{aligned} & \mathbf{cov}(x_{t-1} - (\hat{x}_{t-1}^P)^+) \\ &= \mathbf{cov}(x_{t-1} - \hat{x}_{t-1}^P - J_{t-1}^P (u_{t-1}^E - F_{t-1}^E \hat{x}_{t-1}^P)) \\ &= \mathbf{cov}\left(- (I - J_{t-1}^P F_{t-1}^E) e_{t-1}^P - J_{t-1}^P F_{t-1}^E e_{t-1}^E\right) \\ & \quad + (I - J_{t-1}^P F_{t-1}^E) \tilde{\Sigma}_{t-1}^{(P,E)} (J_{t-1}^P F_{t-1}^E)^\top + (J_{t-1}^P F_{t-1}^E) (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (I - J_{t-1}^P F_{t-1}^E)^\top \\ &= \hat{\Sigma}_{t-1}^P - J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^P - \hat{\Sigma}_{t-1}^P (F_{t-1}^E)^\top (J_{t-1}^P)^\top + J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^P (F_{t-1}^E)^\top (J_{t-1}^P)^\top \\ & \quad + J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^E (F_{t-1}^E)^\top (J_{t-1}^P)^\top + \tilde{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top (J_{t-1}^P)^\top - J_{t-1}^P F_{t-1}^E \tilde{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top (J_{t-1}^P)^\top \\ & \quad + J_{t-1}^P F_{t-1}^E \tilde{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top - J_{t-1}^P F_{t-1}^E (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (F_{t-1}^E)^\top (J_{t-1}^P)^\top. \end{aligned}$$

Note that minimizing $\mathbb{E}[\|(x_{t-1}^P)^+ - x_{t-1}\|^2]$ is equivalent to minimizing $\text{Tr}(\mathbf{cov}(x_{t-1} - (\hat{x}_{t-1}^P)^+))$. Taking the derivative with respect to J_{t-1}^P and setting it to zero, we have

$$\frac{\partial \text{Tr}(\mathbf{cov}(x_{t-1} - (\hat{x}_{t-1}^P)^+))}{\partial J_{t-1}^P} = -2F_{t-1}^E \hat{\Sigma}_{t-1}^P + 2F_{t-1}^E \hat{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top (J_{t-1}^P)^\top + 2F_{t-1}^E (\tilde{\Sigma}_{t-1}^{(P,E)})^\top = 0,$$

which is equivalent to the following equation (since $\hat{\Sigma}_{t-1}^{(P,E)}$ is symmetric by its definition)

$$\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)} = J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^{(P,E)}. \quad (2.28)$$

When $F_{t-1}^E \hat{\Sigma}_{t-1}^{(P,E)} \hat{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top$ is of rank k (which will be shown at the end of the proof), we have

$$J_{t-1}^P = \left(\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)} \right) \hat{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top \left(F_{t-1}^E \hat{\Sigma}_{t-1}^{(P,E)} \hat{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top \right)^{-1}. \quad (2.29)$$

Using the expression in (2.29), we have

$$(\hat{\Sigma}_{t-1}^P)^+ := \mathbf{cov}\left(x_{t-1} - (\hat{x}_{t-1}^P)^+\right) = \hat{\Sigma}_{t-1}^P - (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1}\left(\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)}\right)^\top.$$

Then we have the pre-estimate:

$$(\hat{x}_t^P)^- = A_{t-1}(\hat{x}_{t-1}^P)^+ + B_{t-1}^P u_{t-1}^P + B_{t-1}^E u_{t-1}^E. \quad (2.30)$$

The post-estimate after observing the signal/measure z_t^P at time t is defined as:

$$\hat{x}_t^P = (\hat{x}_t^P)^- + K_t^P \left(z_t^P - H_t^P (\hat{x}_t^P)^-\right), \quad (2.31)$$

with a variance

$$(\hat{\Sigma}_t^P)^- := \mathbb{E}[(\hat{x}_t^P - x_t)(\hat{x}_t^P - x_t)^\top]. \quad (2.32)$$

In the same way as for the derivation of J_{t-1}^P , we can show that the following choice of K_t^P minimizes the quantity $\mathbb{E}[\|x_t - \hat{x}_t^P\|^2]$:

$$K_t^P = (\hat{\Sigma}_{t-1}^P)^- (H_t^P)^\top \left[H_t^P (\hat{\Sigma}_{t-1}^P)^- (H_t^P)^\top + G^P \right]^{-1}. \quad (2.33)$$

The corresponding covariance takes the form:

$$\hat{\Sigma}_t^P := \mathbb{E}[\|x_t - \hat{x}_t^P\|^2] = (I - K_t^P H_t^P)(\hat{\Sigma}_{t-1}^P)^-. \quad (2.34)$$

To update player P 's belief of player E 's state, define similarly to the case (2.18) when $t = 0$,

$$x_t = \hat{x}_t^P - e_t^P = \hat{x}_t^E - e_t^E, \quad (2.35)$$

where e_t^P and e_t^E are the estimation errors from players P and E , respectively. Given that (2.35) is equivalent to the following:

$$\hat{x}_t^P = \hat{x}_t^E + (e_t^P - e_t^E), \quad (2.36)$$

Player P 's posterior distribution for the state estimate \hat{x}_t^E of player E is

$$\hat{x}_t^E \sim \mathcal{N}(\hat{x}_t^P, \hat{\Sigma}_t^{(P,E)}) \quad (2.37)$$

where the estimation error covariance matrix $\hat{\Sigma}_t^{(P,E)}$ is defined as

$$\hat{\Sigma}_t^{(P,E)} = \mathbb{E}[(e_t^E - e_t^P)(e_t^E - e_t^P)^\top] = \hat{\Sigma}_t^P + \hat{\Sigma}_t^E - \tilde{\Sigma}_t^{(P,E)} - \left(\tilde{\Sigma}_t^{(P,E)}\right)^\top, \quad (2.38)$$

in which $\hat{\Sigma}_0^P = W_0^P$, $\hat{\Sigma}_0^E = W_0^E$, $\hat{\Sigma}_t^P = \mathbb{E}[e_t^P(e_t^P)^\top]$, $\hat{\Sigma}_t^E = \mathbb{E}[e_t^E(e_t^E)^\top]$, and $\tilde{\Sigma}_t^{(P,E)} = \mathbb{E}[e_t^P(e_t^E)^\top]$. We will see that $\tilde{\Sigma}_t^{(P,E)}$ satisfies a recursive linear matrix equation which is a Lyapunov equation:

$$\tilde{\Sigma}_t^{(P,E)} = (I - K_t^P H_t^P) \left(A_{t-1} \Delta_{t-1}^{(P,E)} A_{t-1}^\top + \Gamma_{t-1} W \Gamma_{t-1}^\top \right) (I - K_t^E H_t^E)^\top, \quad (2.39)$$

with

$$\begin{aligned} \Delta_{t-1}^{(P,E)} &= (I - J_{t-1}^P F_{t-1}^E) \tilde{\Sigma}_{t-1}^{(P,E)} (I - J_{t-1}^E F_{t-1}^P)^\top + (I - J_{t-1}^P F_{t-1}^E) \hat{\Sigma}_{t-1}^P (J_{t-1}^E F_{t-1}^P)^\top \\ &\quad + J_{t-1}^P F_{t-1}^E (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (J_{t-1}^E F_{t-1}^P)^\top + J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^E (I - J_{t-1}^E F_{t-1}^P)^\top. \end{aligned} \quad (2.40)$$

The equations (2.39) and (2.40) hold since given (2.21d)-(2.21h), we have

$$\hat{x}_t^P = A_{t-1}(\hat{x}_{t-1}^P)^+ + B_{T-1}^P u_{t-1}^P + B_{T-1}^E v_{t-1}^E + K_t^P H_t^P A_{t-1}(x_{t-1} - (\hat{x}_{t-1}^P)^+) + K_t^P H_t^P \Gamma_{t-1} w_{t-1} + K_t^P w_t^P, \quad (2.41)$$

and hence

$$\begin{aligned} e_t^P = \hat{x}_t^P - x_t &= (I - K_t^P H_t^P) A_{t-1}((\hat{x}_{t-1}^P)^+ - x_{t-1}) + (K_t^P H_t^P - I) \Gamma_{t-1} w_{t-1} + K_t^P w_t^P \\ &= (I - K_t^P H_t^P) A_{t-1} \left((I - J_{t-1}^P F_{t-1}^E) e_{t-1}^P + J_{t-1}^P F_{t-1}^E e_{t-1}^E \right) \\ &\quad + (K_t^P H_t^P - I) \Gamma_{t-1} w_{t-1} + K_t^P w_t^P. \end{aligned} \quad (2.42)$$

Similarly, we have

$$e_t^E = (I - K_t^E H_t^E) A_{t-1} \left((I - J_{t-1}^E F_{t-1}^P) e_{t-1}^E + J_{t-1}^E F_{t-1}^P e_{t-1}^P \right) + (K_t^E H_t^E - I) \Gamma_{t-1} w_{t-1} + K_t^E w_t^E. \quad (2.43)$$

Calculating $e_t^P (e_t^E)^\top$ using (2.42) and (2.43) and taking the expectation lead to (2.39) and (2.40).

Now we simplify (2.40) to obtain (2.21j). As for (2.28), we have

$$\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top = J_{t-1}^E F_{t-1}^P \hat{\Sigma}_{t-1}^{(P,E)}. \quad (2.44)$$

By substituting $J_{t-1}^E F_{t-1}^P = (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)(\hat{\Sigma}_{t-1}^{(P,E)})^{-1}$ and $J_{t-1}^P F_{t-1}^E = (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1}$ into (2.40) and by using the fact that $J_{t-1}^E F_{t-1}^P + J_{t-1}^P F_{t-1}^E = I$, we can rewrite $\Delta_{t-1}^{(P,E)}$ as:

$$\begin{aligned} \Delta_{t-1}^{(P,E)} &= J_{t-1}^E F_{t-1}^P \tilde{\Sigma}_{t-1}^{(P,E)} (J_{t-1}^P F_{t-1}^E)^\top + J_{t-1}^E F_{t-1}^P \hat{\Sigma}_{t-1}^P (J_{t-1}^E F_{t-1}^P)^\top \\ &\quad + J_{t-1}^P F_{t-1}^E (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (J_{t-1}^E F_{t-1}^P)^\top + J_{t-1}^P F_{t-1}^E \hat{\Sigma}_{t-1}^E (J_{t-1}^P F_{t-1}^E)^\top \end{aligned} \quad (2.45)$$

$$= (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} \tilde{\Sigma}_{t-1}^{(P,E)} (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top \quad (2.46)$$

$$+ (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} \hat{\Sigma}_{t-1}^P (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)^\top \quad (2.47)$$

$$+ (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)^\top \quad (2.48)$$

$$+ (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} \hat{\Sigma}_{t-1}^E (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top. \quad (2.49)$$

Given the fact that $\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top = \hat{\Sigma}_{t-1}^{(P,E)} - (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})$, we have

$$(2.46) = \tilde{\Sigma}_{t-1}^{(P,E)} (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top \quad (2.50)$$

$$- (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} \tilde{\Sigma}_{t-1}^{(P,E)} (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top. \quad (2.51)$$

Similarly, $\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)} = \hat{\Sigma}_{t-1}^{(P,E)} - (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)$ leads to the following relationship

$$(2.48) = (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - \tilde{\Sigma}_{t-1}^{(P,E)}) \quad (2.52)$$

$$- (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\tilde{\Sigma}_{t-1}^{(P,E)})^\top (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - \tilde{\Sigma}_{t-1}^{(P,E)}). \quad (2.53)$$

Combine (2.51) and (2.49), we have

$$(2.51) + (2.49) = (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - \tilde{\Sigma}_{t-1}^{(P,E)})(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top. \quad (2.54)$$

Combine (2.53) and (2.47), we have

$$(2.53) + (2.47) = (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)(\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - \tilde{\Sigma}_{t-1}^{(P,E)}). \quad (2.55)$$

It is easy to check that $(2.54) + (2.55) = (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)})^\top (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)^\top$. Combining with (2.50) and (2.52), we have

$$\Delta_{t-1}^{(P,E)} = (\hat{\Sigma}_{t-1}^P - \tilde{\Sigma}_{t-1}^{(P,E)}) (\hat{\Sigma}_{t-1}^{(P,E)})^{-1} (\hat{\Sigma}_{t-1}^E - (\tilde{\Sigma}_{t-1}^{(P,E)})^\top)^\top + \tilde{\Sigma}_{t-1}^{(P,E)}. \quad (2.56)$$

Finally we show that $\hat{\Sigma}_{t-1}^{(P,E)}$ is positive definite, which guarantees that $F_{t-1}^E \hat{\Sigma}_{t-1}^{(P,E)} \hat{\Sigma}_{t-1}^{(P,E)} (F_{t-1}^E)^\top$ is of rank k . To see this, we have

$$\hat{\Sigma}_t^{(P,E)} = \mathbb{E} \left[(e_t^E - e_t^P)(e_t^E - e_t^P)^\top \right] = \mathbb{E}[s_{t-1}(s_{t-1})^\top] + K_t^P G^P (K_t^P)^\top + K_t^E G^E (K_t^E)^\top,$$

with s_{t-1} defined as

$$\begin{aligned} s_{t-1} = & (I - K_t^E H_t^E) A_{t-1} \left((I - J_{t-1}^E F_{t-1}^P) e_{t-1}^E + J_{t-1}^E F_{t-1}^P e_{t-1}^P \right) + (K_t^E H_t^E - I) \Gamma_{t-1} w_{t-1} \\ & - (I - K_t^P H_t^P) A_{t-1} \left((I - J_{t-1}^P F_{t-1}^E) e_{t-1}^P + J_{t-1}^P F_{t-1}^E e_{t-1}^E \right) - (K_t^P H_t^P - I) \Gamma_{t-1} w_{t-1}. \end{aligned}$$

It is easy to see that $\mathbb{E}[s_{t-1}(s_{t-1})^\top]$ is positive semi-definite. $K_t^P G^P (K_t^P)^\top$ is positive definite since K_t^P has rank n and G^P is positive definite. Similarly, $K_t^E G^E (K_t^E)^\top$ is also positive definite. \square

2.2 Conditional Expectation and Tower Property

In partially observable game settings, players face the difficult task of incrementally estimating unknown quantities through information filtering, and then using those estimates to make informed decisions. In the linear-quadratic framework, each player i needs to determine

$$\mathbb{E} \left[x_t \mid \mathcal{H}_t^i \right], \text{ and } \mathbb{E} \left[x_t^\top O_t x_t \mid \mathcal{H}_t^i \right], \quad (2.57)$$

for any given matrix $O_t \in \mathbb{R}^{n \times n}$. This requires projection of the unknown quantity into the space spanned by the information filtration \mathcal{H}_t^i . Given that \mathcal{H}_t^i contains information on the opponent's action w_{t-1}^j , the critical challenge boils down to how to utilize this information so that the conditional expectation can be calculated in a valid incremental form to facilitate further analysis and renders the game amenable to solution by dynamic programming.

Theorem 2.4 provides an explicit formula to calculate (2.57) in an incremental format:

$$\mathbb{E} \left[x_t \mid \mathcal{H}_t^i \right] = \hat{x}_t^i, \text{ and } \mathbb{E} \left[x_t^\top O_t x_t \mid \mathcal{H}_t^i \right] = (\hat{x}_t^i)^\top O_t \hat{x}_t^i + \text{Tr} \left(O_t \hat{\Sigma}_t^i \right), \quad (2.58)$$

with \hat{x}_t^i and $\hat{\Sigma}_t^i$ following the explicit recursive formats in (2.21g) and (2.21h), respectively.

Indeed, a similar setup was first studied in [13], in which zero-sum linear-quadratic dynamic games with partial observations and asymmetric information are considered. The players' initial state estimate and their measurements are private information, but each player is able to observe his opponent's past control inputs, so the players' past controls are shared information. However, when the conditional expectation $\mathbb{E}[\cdot \mid \mathcal{H}_t^i]$ is calculated, the recursive format proposed in [13] follows the single-agent Bayes formula (see Equations (11)-(20) in [13] or similarly (2.13a)-(2.13e)) and does not utilize the observable information of the opponent's past controls to improve their state estimation, leading to an incorrect formula and hence the tower property fails to hold, let alone the DPP.

To be mathematically more concrete, we do a sanity check to show that the tower property holds when using the recursive expression in (2.21a)-(2.21k). Namely, from player P 's perspective, we have

$$\mathbb{E} \left[\mathbb{E} \left[x_t^\top Q_t^P x_t \mid \mathcal{H}_t^P \right] \mid \mathcal{H}_{t-1}^P \right] = \mathbb{E} \left[x_t^\top Q_t^P x_t \mid \mathcal{H}_{t-1}^P \right] \quad (2.59)$$

holds when using (2.21a)-(2.21k) to unwind the conditional expectations. To do so, we will first calculate both the LHS and the RHS of (2.59) using (2.21a)-(2.21k), and then match all the terms to prove that the LHS equals the RHS. Finally, we will show that the tower property fails to hold when using the formulas in Equations (11)-(20) of [13] to unwind the conditional expectations.

Calculations using (2.21a)-(2.21k). For the LHS of (2.59), by (2.21g),

$$\begin{aligned}\hat{x}_T^P &= (\hat{x}_T^P)^- + K_T^P \left[z_T^P - H_T^P (\hat{x}_T^P)^- \right] \\ &= (A_{T-1}(\hat{x}_{T-1}^P)^+ + B_{T-1}^P u_{T-1}^P + B_{T-1}^E u_{T-1}^E) + K_T^P w_T^P \\ &\quad + K_T^P H_T^P \left[A_{T-1}(x_{T-1} - (\hat{x}_{T-1}^P)^+) + \Gamma_{T-1} w_{T-1} \right],\end{aligned}\tag{2.60}$$

where (2.60) holds by definition of z_T^P given in (2.3), and (2.21d). We will just consider the case when F_t^P has rank $m < n$ and F_t^E has rank $k < n$ for all $t = 0, 1, \dots, T-1$, as the other cases will follow the same logic. Define $\Pi_{T-1}^P := (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)})(\hat{\Sigma}_{T-1}^{(P,E)})^{-1}$, then (2.28) becomes

$$\Pi_{T-1}^P = J_{T-1}^P F_{T-1}^E.\tag{2.61}$$

We can then rewrite (2.21b) as

$$(\hat{x}_{T-1}^P)^+ = (I - \Pi_{T-1}^P)\hat{x}_{T-1}^P + \Pi_{T-1}^P \hat{x}_{T-1}^E = (I - \Pi_{T-1}^P)\hat{x}_{T-1}^P + \Pi_{T-1}^P(\hat{x}_{T-1}^P - e_{T-1}^P + e_{T-1}^E).$$

Using this equation in (2.60) we obtain

$$\hat{x}_T^P = (A_{T-1} + B_{T-1}^P F_{T-1}^P + B_{T-1}^E F_{T-1}^E)\hat{x}_{T-1}^P + L_1 e_{T-1}^E + L_2 e_{T-1}^P + K_T^P w_T^P + K_T^P H_T^P \Gamma_{T-1} w_{T-1}\tag{2.62}$$

with

$$L_1 := A_{T-1}\Pi_{T-1}^P + B_{T-1}^E F_{T-1}^E - K_T^P H_T^P A_{T-1}\Pi_{T-1}^P,\tag{2.63}$$

$$L_2 := -A_{T-1}\Pi_{T-1}^P - B_{T-1}^E F_{T-1}^E - K_T^P H_T^P A_{T-1}(I - \Pi_{T-1}^P).\tag{2.64}$$

Then substituting (2.62) into the LHS of (2.59), we have

$$\begin{aligned}\mathbb{E} \left[\mathbb{E} \left[x_T^\top Q_T^P x_T \middle| \mathcal{H}_T^P \right] \middle| \mathcal{H}_{T-1}^P \right] &= \mathbb{E} \left[(\hat{x}_T^P)^\top Q_T^P \hat{x}_T^P \middle| \mathcal{H}_{T-1}^P \right] + \text{Tr}(Q_T^P \hat{\Sigma}_T^P) \\ &= (\hat{x}_{T-1}^P)^\top (A_{T-1} + B_{T-1}^P F_{T-1}^P + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (A_{T-1} + B_{T-1}^P F_{T-1}^P + B_{T-1}^E F_{T-1}^E) \hat{x}_{T-1}^P \\ &\quad + \text{Tr}(L_1^\top Q_T^P L_1 \hat{\Sigma}_{T-1}^E) + \text{Tr}(L_2^\top Q_T^P L_2 \hat{\Sigma}_{T-1}^P) + 2 \text{Tr}(L_1^\top Q_T^P L_2 \tilde{\Sigma}_{T-1}^{(P,E)}) \\ &\quad + \text{Tr}((K_T^P)^\top Q_T^P K_T^P G^P) + \text{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) + \text{Tr}(Q_T^P \hat{\Sigma}_T^P).\end{aligned}\tag{2.65}$$

For the RHS of (2.59), we have by expanding x_T directly,

$$\begin{aligned}\mathbb{E} \left[x_T^\top Q_T^P x_T \middle| \mathcal{H}_{T-1}^P \right] &= (\hat{x}_{T-1}^P)^\top (A_{T-1} + B_{T-1}^P F_{T-1}^P + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (A_{T-1} + B_{T-1}^P F_{T-1}^P + B_{T-1}^E F_{T-1}^E) \hat{x}_{T-1}^P \\ &\quad + \text{Tr}(\Gamma_{T-1}^\top Q_T^P \Gamma_{T-1} W) + \text{Tr}((A_{T-1} + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \hat{\Sigma}_{T-1}^P) \\ &\quad + \text{Tr}((F_{T-1}^E)^\top (B_{T-1}^E)^\top Q_T^P B_{T-1}^E F_{T-1}^E \hat{\Sigma}_{T-1}^E) - 2 \text{Tr}((A_{T-1} + B_{T-1}^E F_{T-1}^E)^\top Q_T^P B_{T-1}^E F_{T-1}^E \tilde{\Sigma}_{T-1}^{(E,P)}).\end{aligned}\tag{2.66}$$

The proof that (2.65) is equivalent to (2.66) is deferred to Appendix A.

Calculations using the result in [13]. In [13], the authors used the following recursive formulas which are essentially the same as the single agent case (see Theorem 2.3):

$$x_t \sim \mathcal{N}(\hat{x}_t^i, \hat{\Sigma}_t^i),\tag{2.67}$$

where \hat{x}_t^i and $\hat{\Sigma}_t^i$ are updated according to (2.13a)-(2.13e) Now we use the recursive formula listed in (2.13a)-(2.13e) to calculate the LHS and RHS of (2.59). For the LHS, by direct calculation,

$$\begin{aligned}
& \mathbb{E} \left[(\hat{x}_T^P)^\top Q_T^P \hat{x}_T^P \middle| \mathcal{H}_{T-1}^P \right] + \text{Tr}(Q_T^P \hat{\Sigma}_T^P) \\
&= (\hat{x}_{T-1}^P)^\top A_{T-1}^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P + (\hat{x}_{T-1}^P)^\top (F_{T-1}^P)^\top (B_{T-1}^P)^\top Q_T^P B_{T-1}^P F_{T-1}^P \hat{x}_{T-1}^P \\
&\quad + (F_{T-1}^E \hat{x}_{T-1}^P)^\top ((B_{T-1}^E)^\top Q_T^P B_{T-1}^E) F_{T-1}^E \hat{x}_{T-1}^P + \text{Tr}((F_{T-1}^E)^\top (B_{T-1}^E)^\top Q_T^P B_{T-1}^E F_{T-1}^E \hat{\Sigma}_{T-1}^{(P,E)}) \\
&\quad + 2(F_{T-1}^P \hat{x}_{T-1}^P)^\top (B_{T-1}^P)^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P + 2(F_{T-1}^P \hat{x}_{T-1}^P)^\top (B_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E \hat{x}_{T-1}^P \\
&\quad + 2(F_{T-1}^E \hat{x}_{T-1}^E)^\top (B_{T-1}^E)^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P - 2 \text{Tr}((B_{T-1}^E F_{T-1}^E)^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^{(P,E)}) \\
&\quad + 2 \text{Tr}((B_{T-1}^E F_{T-1}^E)^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} \left(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W \right) \\
&\quad + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P G^P \right) + \text{Tr} \left(A_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^P \right) + \text{Tr}(Q_T^P \hat{\Sigma}_T^P) \quad (2.68)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& \mathbb{E} \left[x_T^\top Q_T^P x_T \middle| \mathcal{H}_{T-1}^P \right] \\
&= (\hat{x}_{T-1}^P)^\top A_{T-1}^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P + (F_{T-1}^P \hat{x}_{T-1}^P)^\top (B_{T-1}^P)^\top Q_T^P B_{T-1}^P (F_{T-1}^P \hat{x}_{T-1}^P) \\
&\quad + (F_{T-1}^E \hat{x}_{T-1}^P)^\top ((B_{T-1}^E)^\top Q_T^P B_{T-1}^E) F_{T-1}^E \hat{x}_{T-1}^P + \text{Tr}(((B_{T-1}^E F_{T-1}^E)^\top Q_T^P B_{T-1}^E F_{T-1}^E) \hat{\Sigma}_{T-1}^{(P,E)}) \\
&\quad + 2u_{T-1}^\top (B_{T-1}^P)^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P + 2u_{T-1}^\top (B_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E \hat{x}_{T-1}^P \\
&\quad + 2(F_{T-1}^E \hat{x}_{T-1}^P)^\top (B_{T-1}^E)^\top Q_T^P A_{T-1} \hat{x}_{T-1}^P + 2 \text{Tr}((B_{T-1}^E F_{T-1}^E)^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) \\
&\quad - 2 \text{Tr}((B_{T-1}^E F_{T-1}^E)^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^{(P,E)}) + \text{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} \left(\Gamma_{T-1}^\top Q_T^P \Gamma_{T-1} W \right) \quad (2.69)
\end{aligned}$$

By routine calculations similar to those used for Step 3 in Appendix A, we see that (2.68) and (2.69) differ from each other by:

$$2 \text{Tr} \left((F_t^E)^\top (B_t^E)^\top Q_{t+1}^P (K_{t+1}^P H_{t+1}^P - I) A_t \left(\hat{\Sigma}_t^{(E,P)} - \hat{\Sigma}_t^P \right) \right). \quad (2.70)$$

Hence the recursive formula (2.13a)-(2.13e) adopted in [13] does not lead to the correct conditional expectation, let alone the tower property and DPP.

3 Equilibrium Solution

Now that we have the information corrections in the updating scheme, we will use these to discuss the DPP and the Nash equilibrium in this section.

3.1 Dynamic Programming Principle

Although both players do not have access to the true state and their controls are related to their state estimates which causes extra correlated randomness, we are still able to derive the individual DPP for the two-player general-sum linear-quadratic Gaussian game under a fixed linear and Markovian strategy from the opponent. This is because the sufficient statistics derived from Theorem 2.4 leads to a valid tower property given in (2.59). This equips us with sufficient tools to prove the DPP.

To start, denote the value function of player i ($i = P, E$), under a fixed strategy $F^j := \{F_t^j\}_{t=0}^{T-1}$ from player j ($j \neq i$) and at any given time $0 \leq t \leq T-1$, as

$$V_t^i(\hat{x}_t^i; F^j) = \min_{\{u_s^i\}_{s=t}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^i x_T + \sum_{s=t}^{T-1} (u_s^i)^\top R_s^i u_s^i + x_t^\top Q_t^i x_t \middle| \mathcal{H}_t^i \right] \quad (3.1)$$

subject to (2.1) and F^j , with the terminal value

$$V_T^i(\hat{x}_T^i; F^j) = V_T^i(\hat{x}_T^i) = \mathbb{E} \left[x_T^\top Q_T^i x_T \mid \mathcal{H}_T^i \right] = (\hat{x}_T^i)^\top Q_T^i \hat{x}_T^i + \text{Tr} \left(Q_T^i \hat{\Sigma}_T^i \right). \quad (3.2)$$

Now we prove the DPP in the two-player general-sum linear-quadratic Gaussian game under partial observations and asymmetric information.

Theorem 3.1 (Dynamic Programming Principle). *For any given time $0 \leq t \leq T-1$, the value function V_t^i for player i ($i = P, E$), under a fixed policy $F^j := \{F_t^j\}_{t=0}^{T-1}$ from player j , satisfies*

$$V_t^i(\hat{x}_t^i; F^j) = \min_{u_t^i} \mathbb{E} \left[x_t^\top Q_t^i x_t + (u_t^i)^\top R_t^i u_t^i + V_{t+1}^i(\hat{x}_{t+1}^i; F^j) \mid \mathcal{H}_t^i \right] \quad (3.3)$$

with $j = P, E$, $i \neq j$, and terminal value $V_T^i(\hat{x}_T^i; F^j)$ given in (3.2).

Proof. We take the perspective of player P and the result for player E follows the same logic. By definition of the value function in (3.1) we have,

$$\begin{aligned} & V_t^P(\hat{x}_t^P; F^E) \\ &= \min_{\{u_s^P\}_{s=t}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^P x_T + \sum_{s=t}^{T-1} (u_s^P)^\top R_s^P u_s^P + x_t^\top Q_t^P x_t \mid \mathcal{H}_t^P \right] \\ &= \min_{u_t^P} \left\{ \mathbb{E} \left[x_t^\top Q_t^P x_t + (u_t^P)^\top R_t^P u_t^P \mid \mathcal{H}_t^P \right] \right. \\ & \quad \left. + \min_{\{u_s^P\}_{s=t+1}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^P x_T + \sum_{s=t+1}^{T-1} \left((u_s^P)^\top R_s^P u_s^P + x_s^\top Q_s^P x_s \right) \mid \mathcal{H}_{t+1}^P \right] \right\} \end{aligned} \quad (3.4)$$

$$\begin{aligned} &= \min_{u_t^P} \left\{ \mathbb{E} \left[x_t^\top Q_t^P x_t + (u_t^P)^\top R_t^P u_t^P \mid \mathcal{H}_t^P \right] \right. \\ & \quad \left. + \min_{\{u_s^P\}_{s=t+1}^{T-1}} \mathbb{E} \left[\mathbb{E} \left[x_T^\top Q_T^P x_T + \sum_{s=t+1}^{T-1} \left((u_s^P)^\top R_s^P u_s^P + x_s^\top Q_s^P x_s \right) \mid \mathcal{H}_{t+1}^P \right] \mid \mathcal{H}_t^P \right] \right\} \end{aligned} \quad (3.5)$$

$$\begin{aligned} &= \min_{u_t^P} \left\{ \mathbb{E} \left[x_t^\top Q_t^P x_t + (u_t^P)^\top R_t^P u_t^P \mid \mathcal{H}_t^P \right] \right. \\ & \quad \left. + \mathbb{E} \left[\min_{\{u_s^P\}_{s=t+1}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^P x_T + \sum_{s=t+1}^{T-1} \left((u_s^P)^\top R_s^P u_s^P + x_s^\top Q_s^P x_s \right) \mid \mathcal{H}_{t+1}^P \right] \mid \mathcal{H}_t^P \right] \right\}, \end{aligned} \quad (3.6)$$

where (3.4) holds since u_t^P is adapted to \mathcal{H}_t^P , (3.5) holds by the tower property (2.59), and (3.6) holds since u_s^P is adapted to \mathcal{H}_s^P ($s \geq t+1$). Finally (3.6) leads to the DPP (3.3) by the definition of V_{t+1}^P . \square

3.2 Nash Equilibrium

In this section, we will show that the Nash equilibrium strategy for the game (2.1)-(2.2)-(2.3)-(2.5)-(2.7) is related to the solution of a coupled Riccati system. Assumption 3.2 is the existence and uniqueness of this solution and we provide a sufficient condition for this assumption in Remark 3.3.

Assumption 3.2. *There exists a unique solution set $F^{P*} := \{F_t^{P*}\}_{t=0}^{T-1}$ with $F_t^{P*} \in \mathcal{A}^P$ and $F^{E*} := \{F_t^{E*}\}_{t=0}^{T-1}$ with $F_t^{E*} \in \mathcal{A}^E$ to the following set of linear matrix equations:*

$$F_t^{P*} = -(R_t^P + (B_t^P)^\top U_{t+1}^{P*} B_t^P)^{-1} ((B_t^P)^\top U_{t+1}^{P*} (A_t + B_t^E F_t^{E*})), \quad (3.7)$$

$$F_t^{E*} = -(R_t^E + (B_t^E)^\top U_{t+1}^{E*} B_t^E)^{-1} ((B_t^E)^\top U_{t+1}^{E*} (A_t + B_t^P F_t^{P*})), \quad (3.8)$$

where $\{U_t^{P*}\}_{t=0}^T$ and $\{U_t^{E*}\}_{t=0}^T$ are obtained recursively backwards from

$$U_t^{P*} = Q_t^P + (F_t^{P*})^\top R_t^P F_t^{P*} + (A_t + B_t^P F_t^{P*} + B_t^E F_t^{E*})^\top U_{t+1}^{P*} (A_t + B_t^P F_t^{P*} + B_t^E F_t^{E*}), \quad (3.9)$$

$$U_t^{E*} = Q_t^E + (F_t^{E*})^\top R_t^E F_t^{E*} + (A_t + B_t^P F_t^{P*} + B_t^E F_t^{E*})^\top U_{t+1}^{E*} (A_t + B_t^P F_t^{P*} + B_t^E F_t^{E*}) \quad (3.10)$$

with terminal conditions $U_T^{i*} = Q_T^i$ for $i = P, E$.

Remark 3.3. A sufficient condition for the unique solvability of (3.9)-(3.10) is the invertibility of the block matrix Φ_t , $t = 0, 1, \dots, T-1$, with the ii -th block given by $R_t^i + (B_t^i)^\top U_{t+1}^{i*} B_t^i$ and the ij -th block given by $(B_t^i)^\top U_{t+1}^{i*} B_t^j$, where $i, j = P, E$ and $j \neq i$. See Remark 6.5 in [2].

Using the DPP formula in Theorem 3.1, we have the following result for the Nash equilibrium strategy and the corresponding value function for the game (2.1)-(2.2)-(2.3)-(2.5)-(2.7).

Theorem 3.4. Suppose Assumptions 2.1 and 3.2 hold. We also assume that both players are applying linear strategies. Then the unique Nash equilibrium policy can be expressed as for $i = P, E$

$$u_t^{i*}(\hat{x}_t^i) = F_t^{i*} \hat{x}_t^i, \quad (3.11)$$

with F_t^{P*} and F_t^{E*} given in (3.7) and (3.8). The corresponding optimal value function of player i is quadratic ($0 \leq t \leq T$):

$$V_t^i(\hat{x}_t^i; F^{j*}) = (\hat{x}_t^i)^\top U_t^{i*} \hat{x}_t^i + c_t^{i*}, \quad (3.12)$$

where $j = P, E$ and $j \neq i$, the matrices $U_t^{P*}, U_t^{E*} \in \mathbb{R}^{n \times n}$ are given in (3.9) and (3.10), and the scalars $c_t^{P*}, c_t^{E*} \in \mathbb{R}$ are given by

$$\begin{aligned} c_t^{i*} &= c_{t+1}^{i*} + \text{Tr} \left(Q_t^i \hat{\Sigma}_t^i \right) - \text{Tr} \left(U_{t+1}^{i*} \hat{\Sigma}_{t+1}^i \right) + \text{Tr} \left((A_t + B_t^j F_t^{j*})^\top U_{t+1}^{i*} (A_t + B_t^j F_t^{j*}) \hat{\Sigma}_t^i \right) \\ &\quad + \text{Tr} \left((F_t^{j*})^\top (B_t^j)^\top U_{t+1}^{i*} B_t^j F_t^{j*} \hat{\Sigma}_t^j \right) - 2 \text{Tr} \left((A_t + B_t^j F_t^{j*})^\top U_{t+1}^{i*} B_t^j F_t^{j*} \hat{\Sigma}_t^{(j,i)} \right) \\ &\quad + \text{Tr}(\Gamma_t^\top U_{t+1}^{i*} \Gamma_t W). \end{aligned} \quad (3.13)$$

The terminal condition for player i is $c_T^{i*} = \text{Tr} \left(Q_T^i \hat{\Sigma}_T^i \right)$.

Remark 3.5 (Discussion of linear policies). 1. In the partially observable setting, it is widely recognized that the existence of a Nash equilibrium is not guaranteed if a more general class of policies is considered, as players can mislead their opponents by disclosing false intentions [5, 20].

2. We note that the optimal policies F_t^{P*} and F_t^{E*} given in (3.7) and (3.8), and the Riccati equations given in (3.9) and (3.10), are the same as the optimal policies and Riccati equations in the case of full observation ([2, Corollary 6.4]). However, the linear-quadratic Gaussian game under partial observation (defined in (2.1)-(2.2)-(2.3)-(2.5)-(2.7)) differs from the linear-quadratic game with *full information* in [2, Corollary 6.4] in the sense that the Nash equilibrium strategy is linear in the *state estimate* rather than the *true state*, and the scalars c_t^{P*} and c_t^{E*} in the value function involve more terms due to the errors in state estimation.

Proof. We prove the theorem by backward induction. We take the perspective of player P and let player E use the linear strategy $F^{E*} = \{F_t^{E*}\}_{t=0}^{T-1}$ defined in (3.8). At time T , (3.12) holds by the terminal condition given in (3.2). At time $T-1$, by Theorem 3.1, we have the DPP for player P :

$$V_{T-1}^P(\hat{x}_{T-1}^P; F^{E*}) = \min_{u_{T-1}^P} \mathbb{E} \left[x_{T-1}^\top Q_{T-1}^P x_{T-1} + (u_{T-1}^P)^\top R_{T-1}^P u_{T-1}^P + V_T^P(\hat{x}_T^P; F^{E*}) \middle| \mathcal{H}_{T-1}^P \right],$$

and by (2.62),

$$\hat{x}_T^P = (A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \hat{x}_{T-1}^P + B_{T-1}^P u_{T-1}^P + L_{T-1}^1 e_{T-1}^E + L_2 e_{T-1}^P + K_T^P w_T^P + K_T^P H_T^P \Gamma_{T-1} w_{T-1},$$

with L_{T-1}^1 and L_{T-1}^2 defined as $L_{T-1}^1 = A_{T-1} \Pi_{T-1}^P + B_{T-1}^E F_{T-1}^{E*} - K_T^P H_T^P A_{T-1} \Pi_{T-1}^P$ and $L_{T-1}^2 = -A_{T-1} \Pi_{T-1}^P - B_{T-1}^E F_{T-1}^{E*} - K_T^P H_T^P A_{T-1} (I - \Pi_{T-1}^P)$, where Π_{T-1}^P is defined as $\Pi_{T-1}^P = (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) (\hat{\Sigma}_{T-1}^{(P,E)})^{-1}$. Hence

$$\begin{aligned} V_{T-1}^P(\hat{x}_{T-1}^P; F^{E*}) &= \min_{u_{T-1}^P} \left\{ (u_{T-1}^P)^\top R_{T-1}^P u_{T-1}^P + (\hat{x}_{T-1}^P)^\top Q_T^P \hat{x}_{T-1}^P + \text{Tr} \left(Q_{T-1}^P \hat{\Sigma}_{T-1}^P \right) \right. \\ &\quad \left. + \mathbb{E} \left[V_T^P((A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \hat{x}_{T-1}^P + B_{T-1}^P u_{T-1}^P + L_{T-1}^1 e_{T-1}^E + L_{T-1}^2 e_{T-1}^P \right. \right. \\ &\quad \left. \left. + K_T^P w_T^P + K_T^P H_T^P \Gamma_{T-1} w_{T-1}; F^{E*} \right) \middle| \mathcal{H}_{T-1}^P \right] \right\}, \end{aligned} \quad (3.14)$$

Since $V_T^P(\hat{x}_T^P) = (\hat{x}_T^P)^\top Q_T^P \hat{x}_T^P + \text{Tr}(Q_T^P \hat{\Sigma}_T^P)$, we have

$$\begin{aligned} V_{T-1}^P(\hat{x}_{T-1}^P; F^{E*}) &= \min_{u_{T-1}^P} \left\{ (u_{T-1}^P)^\top R_{T-1}^P u_{T-1}^P + (\hat{x}_{T-1}^P)^\top Q_T^P \hat{x}_{T-1}^P + \text{Tr} (Q_{T-1}^P \hat{\Sigma}_{T-1}^P) \right. \\ &\quad \left. + \text{Tr} (Q_T^P \hat{\Sigma}_T^P) + \mathbb{E} \left[((A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \hat{x}_{T-1}^P + B_{T-1}^P u_{T-1}^P + L_{T-1}^1 e_{T-1}^E + L_{T-1}^2 e_{T-1}^P \right. \right. \\ &\quad \left. \left. + K_T^P w_T^P + K_T^P H_T^P \Gamma_{T-1} w_{T-1})^\top Q_T^P ((A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \hat{x}_{T-1}^P + B_{T-1}^P u_{T-1}^P \right. \right. \\ &\quad \left. \left. + L_{T-1}^1 e_{T-1}^E + L_{T-1}^2 e_{T-1}^P + K_T^P w_T^P + K_T^P H_T^P \Gamma_{T-1} w_{T-1}) \middle| \mathcal{H}_{T-1}^P \right] \right\}. \end{aligned} \quad (3.15)$$

Expanding terms in the expectation, (3.15) becomes

$$\begin{aligned} &V_{T-1}^P(\hat{x}_{T-1}^P; F^{E*}) \\ &= \min_{u_{T-1}^P} \left\{ (u_{T-1}^P)^\top (R_{T-1}^P + (B_{T-1}^P)^\top Q_T^P B_{T-1}^P) u_{T-1}^P + 2(\hat{x}_{T-1}^P)^\top (A_{T-1} + B_{T-1}^E F_{T-1}^{E*})^\top Q_T^P B_{T-1}^P u_{T-1}^P \right\} \\ &\quad + (\hat{x}_{T-1}^P)^\top \left(Q_{T-1}^P + (A_{T-1} + B_{T-1}^E F_{T-1}^{E*})^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \right) \hat{x}_{T-1}^P + \text{Tr} (Q_T^P \hat{\Sigma}_T^P) \\ &\quad + \text{Tr} (Q_{T-1}^P \hat{\Sigma}_{T-1}^P) + \text{Tr}((L_{T-1}^1)^\top Q_T^P L_{T-1}^1 \hat{\Sigma}_{T-1}^E) + \text{Tr}((L_{T-1}^2)^\top Q_T^P L_{T-1}^2 \hat{\Sigma}_{T-1}^P) \\ &\quad + 2 \text{Tr}((L_{T-1}^1)^\top Q_T^P L_{T-1}^2 \tilde{\Sigma}_{T-1}^{(P,E)}) + \text{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) \\ &\quad + \text{Tr}((K_T^P)^\top Q_T^P K_T^P G^P). \end{aligned} \quad (3.16)$$

We note that all the constant terms are independent of u_{T-1}^P , since $\hat{\Sigma}_T^P$, $\tilde{\Sigma}_{T-1}^{(P,E)}$, and $\hat{\Sigma}_{T-1}^P$ are independent of the policy F_{T-1}^P . Thus these constant terms will not be involved in the minimization problem. Applying the first order condition to the minimization part in (3.16) leads to

$$u_{T-1}^{P*} = -(R_{T-1}^P + (B_{T-1}^P)^\top Q_T^P B_{T-1}^P)^{-1} (B_{T-1}^P)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^{E*}) \hat{x}_{T-1}^P = F_{T-1}^{P*} \hat{x}_{T-1}^P.$$

Similarly, we can derive the optimal policy of the player E when fixing player P 's strategy F_{T-1}^{P*} :

$$u_{T-1}^{E*} = -(R_{T-1}^E + (B_{T-1}^E)^\top Q_T^E B_{T-1}^E)^{-1} (B_{T-1}^E)^\top Q_T^E (A_{T-1} + B_{T-1}^P F_{T-1}^{P*}) \hat{x}_{T-1}^E = F_{T-1}^{E*} \hat{x}_{T-1}^E.$$

Substituting $u_{T-1}^{P*} = F_{T-1}^{P*} \hat{x}_{T-1}^P$ into (3.16) we obtain the optimal value function given as

$$\begin{aligned} V_{T-1}^P(\hat{x}_{T-1}^P; F^{E*}) &= (\hat{x}_{T-1}^P)^\top \left(Q_{T-1}^P + (F_{T-1}^{P*})^\top R_{T-1}^P F_{T-1}^{P*} + (A_{T-1} + B_{T-1}^P F_{T-1}^{P*} + B_{T-1}^E F_{T-1}^{E*})^\top \right. \\ &\quad \left. Q_T^P (A_{T-1} + B_{T-1}^P F_{T-1}^{P*} + B_{T-1}^E F_{T-1}^{E*}) \right) \hat{x}_{T-1}^P + \text{Tr} (Q_{T-1}^P \hat{\Sigma}_{T-1}^P) \\ &\quad + \text{Tr} (Q_T^P \hat{\Sigma}_T^P) + \text{Tr}((L_{T-1}^1)^\top Q_T^P L_{T-1}^1 \hat{\Sigma}_{T-1}^E) + \text{Tr}((L_{T-1}^2)^\top Q_T^P L_{T-1}^2 \hat{\Sigma}_{T-1}^P) \end{aligned}$$

$$\begin{aligned}
& +2 \operatorname{Tr}((L_{T-1}^1)^\top Q_T^P L_{T-1}^2 \tilde{\Sigma}_{T-1}^{(P,E)}) + \operatorname{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) \\
& + \operatorname{Tr}((K_T^P)^\top Q_T^P K_T^P G^P)
\end{aligned} \tag{3.17}$$

$$= (\hat{x}_{T-1}^P)^\top U_{T-1}^{P*} (\hat{x}_{T-1}^P)^\top + c_{T-1}^{P*}. \tag{3.18}$$

where (3.18) holds by a similar calculation to that proving (2.65) is equivalent to (2.66) in Section 2.2. Similarly we can also show that (3.12) holds for player E at time $T-1$.

Now assume that (3.11)-(3.12) holds at all $s \geq t+1$. Then we have

$$V_{t+1}^P(\hat{x}_{t+1}^P; F^{E*}) = (\hat{x}_{t+1}^P)^\top U_{t+1}^{P*} \hat{x}_{t+1}^P + c_{t+1}^{P*}. \tag{3.19}$$

At time t , recall that Π_t^P is defined as $\Pi_t^P = (\hat{\Sigma}_t^P - \tilde{\Sigma}_t^{(P,E)})(\hat{\Sigma}_t^{(P,E)})^{-1}$. We further define L_t^1 and L_t^2 as $L_t^1 = A_t \Pi_t^P + B_t^E F_t^{E*} - K_{t+1}^P H_{t+1}^P A_t \Pi_t^P$ and $L_t^2 = -A_t \Pi_t^P - B_t^E F_t^{E*} - K_{t+1}^P H_{t+1}^P A_t (I - \Pi_t^P)$. Then similarly to (2.62) we have

$$\hat{x}_{t+1}^P = (A_t + B_t^E F_t^{E*}) \hat{x}_t^P + B_t^P u_t^P + L_t^1 e_t^E + L_t^2 e_t^P + K_{t+1}^P w_{t+1}^P + K_{t+1}^P H_{t+1}^P \Gamma_t w_t, \tag{3.20}$$

We apply the DPP again at time t , then by (3.19) and (3.20) we have

$$\begin{aligned}
& V_t^P(\hat{x}_t^P; F^{E*}) \\
= & \min_{u_t^P} \left\{ (u_t^P)^\top R_t^P u_t^P + (\hat{x}_t^P)^\top Q_t^P \hat{x}_t^P + \operatorname{Tr}(Q_t^P \hat{\Sigma}_t^P) \right. \\
& + \mathbb{E} \left[((A_t + B_t^E F_t^{E*}) \hat{x}_t^P + B_t^P u_t^P + L_t^1 e_t^E + L_t^2 e_t^P + K_{t+1}^P w_{t+1}^P + K_{t+1}^P H_{t+1}^P \Gamma_t w_t)^\top \cdot U_{t+1}^{P*} \right. \\
& \left. \left. ((A_t + B_t^E F_t^{E*}) \hat{x}_t^P + B_t^P u_t^P + L_t^1 e_t^E + L_t^2 e_t^P + K_{t+1}^P w_{t+1}^P + K_{t+1}^P H_{t+1}^P \Gamma_t w_t) + c_{t+1}^{P*} | \mathcal{H}_t^P \right] \right\}.
\end{aligned}$$

Expanding the terms in the expectation we obtain

$$\begin{aligned}
& V_t^P(\hat{x}_t^P; F^{E*}) \\
= & \min_{u_t^P} \left\{ (u_t^P)^\top (R_t^P + (B_t^P)^\top U_{t+1}^{P*} B_t^P) u_t^P + 2(\hat{x}_t^P)^\top (A_t + B_t^E F_t^{E*})^\top U_{t+1}^{P*} B_t^P u_t^P \right\} \\
& + c_{t+1}^{P*} + (\hat{x}_t^P)^\top (Q_t^P + (A_t + B_t^E F_t^{E*})^\top U_{t+1}^{P*} (A_t + B_t^E F_t^{E*})) \hat{x}_t^P + \operatorname{Tr}(Q_t^P \hat{\Sigma}_t^P) \\
& + \operatorname{Tr}((L_t^1)^\top U_{t+1}^{P*} L_t^1 \hat{\Sigma}_t^E) + \operatorname{Tr}((L_t^2)^\top U_{t+1}^{P*} L_t^2 \hat{\Sigma}_t^P) + 2 \operatorname{Tr}((L_t^1)^\top U_{t+1}^{P*} L_t^2 \tilde{\Sigma}_t^{(P,E)}) \\
& + \operatorname{Tr}(\Gamma_t^\top (H_{t+1}^P)^\top (K_{t+1}^P)^\top U_{t+1}^{P*} K_{t+1}^P H_{t+1}^P \Gamma_t W) + \operatorname{Tr}((K_{t+1}^P)^\top U_{t+1}^{P*} K_{t+1}^P G^P)
\end{aligned} \tag{3.21}$$

We note that all the constant terms including the accumulated sum c_{t+1}^{P*} are independent of u_t^P , since $\hat{\Sigma}_s^P$, $\hat{\Sigma}_s^E$, and $\tilde{\Sigma}_s^{(P,E)}$ ($s = t, \dots, T$) are independent of the sequence $\{F_s^P\}_{s=0}^t$.

We can apply the first-order condition to obtain the following optimal response:

$$u_t^{P*} = -(R_t^P + (B_t^P)^\top U_{t+1}^{P*} B_t^P)^{-1} (B_t^P)^\top U_{t+1}^{P*} (A_t + B_t^E F_t^{E*}) \hat{x}_t^P = F_t^{P*} \hat{x}_t^P. \tag{3.22}$$

Similarly, player E minimizes his value function to find his optimal response to player P 's strategy $\hat{F}_t^{P*} \hat{x}_t^P$. We can show that the optimal strategy u_t^{E*} of player E is given by

$$u_t^{E*} = -(R_t^E + (B_t^E)^\top U_{t+1}^{E*} B_t^E)^{-1} (B_t^E)^\top U_{t+1}^{E*} (A_t + B_t^P F_t^{P*}) \hat{x}_t^E = F_t^{E*} \hat{x}_t^E. \tag{3.23}$$

Plugging $u_t^{P*} = F_t^{P*} \hat{x}_t^P$ into (3.21) and after manipulations similar to those in the proof that (2.65) is equivalent to (2.66) in Section 2.2, we can rewrite the value function as

$$V_t^P(\hat{x}_t^P; F^{E*}) = (\hat{x}_t^P)^\top U_t^{P*} \hat{x}_t^P + c_t^{P*},$$

with U_t^{P*} and c_t^{P*} given in (3.9) and (3.13). Similarly, we can show that (3.13) also holds for player E . Therefore by backward induction, the statements hold for all $t = 0, 1, \dots, T$. \square

4 A Mixed Partially and Fully Observable Setting

In this section, we consider a more general setting for games with two players, P and E . Now part of the state process is fully observable and part of the state process is partially observable. The joint dynamics $x_t \in \mathbb{R}^n$ takes a linear form ($0 \leq t \leq T-1$):

$$x_{t+1} := \begin{pmatrix} x_{t+1}^{(1)} \\ x_{t+1}^{(2)} \end{pmatrix} = A_t \begin{pmatrix} x_t^{(1)} \\ x_t^{(2)} \end{pmatrix} + B_t^P u_t^P + B_t^E u_t^E + \Gamma_t w_t, \quad (4.1)$$

with initial value $x_0 = (x_0^{(1)}, x_0^{(2)})^\top$, and the controls of P and E are $u_t^P \in \mathbb{R}^m$ and $u_t^E \in \mathbb{R}^k$, respectively. Here, for each t , the noise $w_t \in \mathbb{R}^d$ is an i.i.d. sample from $\mathcal{N}(0, W)$ with $W \in \mathbb{R}^{d \times d}$ and we have the model parameters $A_t \in \mathbb{R}^{n \times n}$, $B_t^P \in \mathbb{R}^{n \times m}$, $B_t^E \in \mathbb{R}^{n \times k}$, and $\Gamma_t \in \mathbb{R}^{n \times d}$. We assume that $x_t^{(1)} \in \mathbb{R}^{n_1}$ is the partially observable part and $x_t^{(2)} \in \mathbb{R}^{n_2}$ is the fully observable part with $n = n_1 + n_2$.

Information Structure. At the time $t = 0$, player P observes $x_0^{(2)}$ and believes that $x_0^{(1)}$ is drawn from a Gaussian distribution $x_0^{(1)} \sim \mathcal{N}(\hat{x}_0^{P,(1)}, W_0^P)$, and thereafter player P observes part of the state $x_t^{(2)} \in \mathbb{R}^{n_2}$ and the noisy state signal $z_t^P \in \mathbb{R}^p$:

$$z_{t+1}^P = H_{t+1}^P x_{t+1}^{(1)} + w_{t+1}^P, \quad w_{t+1}^P \sim \mathcal{N}(0, G^P), \quad t = 0, 1, \dots, T-1, \quad (4.2)$$

with $\{w_t^P\}_{t=0}^{T-1}$ a sequence of i.i.d. random variables. Here $G^P \in \mathbb{R}^{p \times p}$ and $H_{t+1}^P \in \mathbb{R}^{p \times n_1}$. Similarly, player E observes $x_0^{(2)}$ and believes that $x_0^{(1)}$ is drawn from a Gaussian distribution $x_0^{(1)} \sim \mathcal{N}(\hat{x}_0^{E,(1)}, W_0^E)$. Then player E observes part of the state $x_t^{(2)} \in \mathbb{R}^{n_2}$ and the noisy state signal $z_t^E \in \mathbb{R}^q$:

$$z_{t+1}^E = H_{t+1}^E x_{t+1}^{(1)} + w_{t+1}^E, \quad w_{t+1}^E \sim \mathcal{N}(0, G^E), \quad t = 0, 1, \dots, T-1. \quad (4.3)$$

with $\{w_t^E\}_{t=0}^{T-1}$ a sequence of i.i.d. random variables. For simplicity we assume that $\{w_t^E\}_{t=0}^{T-1}$ are independent from $\{w_t^P\}_{t=0}^{T-1}$. In addition we have $G^E \in \mathbb{R}^{q \times q}$ and $H_{t+1}^E \in \mathbb{R}^{q \times n_1}$.

Both players make their decisions based on the public and private information available to them. We write $\mathcal{Z}_t^P = \{z_s^P\}_{s=1}^t$ and $\mathcal{Z}_t^E = \{z_s^E\}_{s=1}^t$ for the private signals players P and E receive up to time t ($1 \leq t \leq T$), respectively. Let $\mathcal{U}_t^P = \{u_s^P\}_{s=1}^t$ and $\mathcal{U}_t^E = \{u_s^E\}_{s=1}^t$ denote the control history from the buyer and seller up to time t , respectively. Also let $\mathcal{X}_t := \{x_s^{(2)}\}_{s=0}^t$ be the public information that is available to both players.

We assume \mathcal{H}_t^P is the information (or history) available to player P and \mathcal{H}_t^E is the information available to player E for them to make decisions at time t , where \mathcal{H}_t^P and \mathcal{H}_t^E follow:

$$\mathcal{H}_t^P = \{\hat{x}_0^P, W_0^P, W_0^E\} \cup \mathcal{Z}_t^P \cup \mathcal{X}_t \cup \mathcal{U}_{t-1}^P \cup \mathcal{U}_{t-1}^E, \quad \mathcal{H}_t^E = \{\hat{x}_0^E, W_0^P, W_0^E\} \cup \mathcal{Z}_t^E \cup \mathcal{X}_t \cup \mathcal{U}_{t-1}^P \cup \mathcal{U}_{t-1}^E. \quad (4.4)$$

Note that the covariance matrices $\{W_0^P, W_0^E\}$ are known to both players.

Cost Function. Each player i ($i = P, E$) strives to minimize their own cost function:

$$\min_{\{u_t^i\}_{t=0}^{T-1}} J^i(\hat{x}_0^{i,(1)}, x_0^{(2)}) := \min_{\{u_t^i\}_{t=0}^{T-1}} \mathbb{E} \left[x_T^\top Q_T^i x_T + \sum_{t=0}^{T-1} \left(x_t^\top Q_t^i x_t + (u_t^i)^\top R_t^i u_t^i \right) \middle| \mathcal{H}_0^i \right], \quad (4.5)$$

with cost parameters $Q_t^P, Q_t^E \in \mathbb{R}^{n \times n}$, $R_t^P \in \mathbb{R}^{m \times m}$ and $R_t^E \in \mathbb{R}^{k \times k}$.

Rewrite the earlier model as $A_t = \begin{pmatrix} A_t^{(1,1)} & A_t^{(1,2)} \\ A_t^{(2,1)} & A_t^{(2,2)} \end{pmatrix}$ with $A_t^{(1,1)} \in \mathbb{R}^{n_1 \times n_1}$, $A_t^{(1,2)} \in \mathbb{R}^{n_1 \times n_2}$, $A_t^{(2,1)} \in \mathbb{R}^{n_2 \times n_1}$ and $A_t^{(2,2)} \in \mathbb{R}^{n_2 \times n_2}$. Similarly, rewrite $B_t^P = (B_t^{P,(1)}, B_t^{P,(2)})^\top$ with $B_t^{P,(1)} \in \mathbb{R}^{n_1 \times m}$ and

$B_t^{P,(2)} \in \mathbb{R}^{n_2 \times m}$, and $B_t^E = (B_t^{E,(1)}, B_t^{E,(2)})^\top$ with $B_t^{E,(1)} \in \mathbb{R}^{n_1 \times k}$, $B_t^{E,(2)} \in \mathbb{R}^{n_2 \times k}$, and $\Gamma_t = (\Gamma_t^{(1)}, \Gamma_t^{(2)})^\top$ with $\Gamma_t^{(1)} \in \mathbb{R}^{n_1 \times d}$ and $\Gamma_t^{(2)} \in \mathbb{R}^{n_2 \times d}$. For the cost parameters, $Q_t^i = \begin{pmatrix} Q_t^{i,(1,1)} & Q_t^{i,(1,2)} \\ (Q_t^{i,(1,2)})^\top & Q_t^{i,(2,2)} \end{pmatrix}$ with $Q_t^{i,(1,1)} \in \mathbb{R}^{n_1 \times n_1}$, $Q_t^{i,(1,2)} \in \mathbb{R}^{n_1 \times n_2}$, and $Q_t^{i,(2,2)} \in \mathbb{R}^{n_2 \times n_2}$ for $i = P, E$.

For the mixed case we make the following assumptions on the parameters, initial state, and noise.

Assumption 4.1 ([Mixed Setting] Parameters, Initial State, and Noise). *For $i = P, E$,*

1. $\{w_t\}_{t=0}^{T-1}$ and $\{w_t^i\}_{t=1}^{T-1}$ are zero-mean, i.i.d. Gaussian random variables that are independent from x_0 and each other and such that $\mathbb{E}[w_t w_t^\top] = W$ and $\mathbb{E}[w_t^i (w_t^i)^\top] = G^i$ are positive definite;
2. Both matrices $H_{t+1}^P \in \mathbb{R}^{p \times n_1}$ and $H_{t+1}^E \in \mathbb{R}^{q \times n_1}$ have rank n_1 for $t = 0, \dots, T-1$.
3. The matrices $\Gamma_t^{(1)} W (\Gamma_t^{(1)})^\top$ are non-singular for $t = 1, \dots, T$;
4. The cost matrices Q_t^i , for $t = 0, 1, \dots, T$ are positive semi-definite, and R_t^i for $t = 0, 1, \dots, T-1$ are positive definite.

We now give the main results and omit the proofs as they follow naturally by applying the ideas in Sections 2 and 3 to the partially observable part of the state process.

Theorem 4.2 (Sufficient Statistics in Two-player Games). *Assume that both players are applying linear strategies in that $u_t^P = F_t^{P,(1)} \mathbb{E}[x_t^{(1)} | \mathcal{H}_t^P] + F_t^{P,(2)} x_t^{(2)}$ and $u_t^E = F_t^{E,(1)} \mathbb{E}[x_t^{(1)} | \mathcal{H}_t^E] + F_t^{E,(2)} x_t^{(2)}$ for some matrices $F_t^{P,(1)} \in \mathbb{R}^{m \times n_1}$ of rank $\min(m, n_1)$, $F_t^{P,(2)} \in \mathbb{R}^{m \times n_2}$, $F_t^{E,(1)} \in \mathbb{R}^{k \times n_1}$ of rank $\min(k, n_1)$, and $F_t^{E,(2)} \in \mathbb{R}^{k \times n_2}$. The sufficient statistic of player i for $i = P, E$ at decision time $t = 0$ is $x_0 \sim N(\hat{x}_0^i, W_0^i)$. For time $1 \leq t \leq T-1$, the distribution of $x_t^{(1)}$ as calculated by player i , conditioning on the private information available to him at time t , is given by*

$$x_t^{(1)} \sim \mathcal{N}(\hat{x}_t^{i,(1)}, \hat{\Sigma}_t^i), \quad (4.6)$$

where, for $j \neq i$,

$$J_{t-1}^i = \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right) \hat{\Sigma}_{t-1}^{(i,j)} (Y_{t-1}^{j,(1)})^\top \left(Y_{t-1}^{j,(1)} \hat{\Sigma}_{t-1}^{(i,j)} \hat{\Sigma}_{t-1}^{(i,j)} (Y_{t-1}^{j,(1)})^\top \right)^{-1}, \quad (4.7a)$$

$$(\hat{x}_{t-1}^{i,(1)})^+ = \hat{x}_{t-1}^{i,(1)} + J_{t-1}^i \left(y_{t-1}^j - Y_{t-1}^{j,(1)} \hat{x}_{t-1}^{i,(1)} \right), \quad (4.7b)$$

$$(\hat{\Sigma}_{t-1}^i)^+ = \hat{\Sigma}_{t-1}^i - \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right) \left(\hat{\Sigma}_{t-1}^{(i,j)} \right)^{-1} \left(\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)} \right)^\top, \quad (4.7c)$$

$$(\hat{x}_t^{i,(1)})^- = A_{t-1}^{(1,1)} (\hat{x}_{t-1}^{i,(1)})^+ + A_{t-1}^{(1,2)} x_{t-1}^{(2)} + B_{t-1}^{P,(1)} u_{t-1}^P + B_{t-1}^{E,(1)} u_{t-1}^E, \quad (4.7d)$$

$$(\hat{\Sigma}_t^i)^- = A_{t-1}^{(1,1)} (\hat{\Sigma}_{t-1}^i)^+ (A_{t-1}^{(1,1)})^\top + \Gamma_{t-1}^{(1)} W (\Gamma_{t-1}^{(1)})^\top, \quad (4.7e)$$

$$K_t^i = (\hat{\Sigma}_t^i)^- (H_t^i)^\top \left[H_t^i (\hat{\Sigma}_t^i)^- (H_t^i)^\top + G^i \right]^{-1}, \quad (4.7f)$$

$$\hat{x}_t^{i,(1)} = (\hat{x}_t^{i,(1)})^- + K_t^i \left[z_t^i - H_t^i (\hat{x}_t^{i,(1)})^- \right], \quad (4.7g)$$

$$\hat{\Sigma}_t^i = (I - K_t^i H_t^i) (\hat{\Sigma}_t^i)^-, \quad (4.7h)$$

$$\tilde{\Sigma}_t^{(i,j)} = (I - K_t^i H_t^i) \left(A_{t-1}^{(1,1)} \Delta_{t-1}^{(i,j)} (A_{t-1}^{(1,1)})^\top + \Gamma_{t-1}^{(1)} W (\Gamma_{t-1}^{(1)})^\top \right) (I - K_t^j H_t^j)^\top, \quad (4.7i)$$

$$\Delta_{t-1}^{(i,j)} = (\hat{\Sigma}_{t-1}^i - \tilde{\Sigma}_{t-1}^{(i,j)}) (\hat{\Sigma}_{t-1}^{(i,j)})^{-1} (\hat{\Sigma}_{t-1}^j - \tilde{\Sigma}_{t-1}^{(j,i)})^\top + \tilde{\Sigma}_{t-1}^{(i,j)}, \quad (4.7j)$$

$$\hat{\Sigma}_t^{(i,j)} = \hat{\Sigma}_t^i + \hat{\Sigma}_t^j - \tilde{\Sigma}_t^{(i,j)} - \left(\tilde{\Sigma}_t^{(i,j)} \right)^\top, \quad (4.7k)$$

where $\widehat{\Sigma}_{t-1}^{(i,j)}$ is positive definite. The values of $Y_t^{P,(1)} \in \mathbb{R}^{m \times n_1}$, $Y_t^{E,(1)} \in \mathbb{R}^{k \times n_1}$ and y_t^P, y_t^E depend on the ranks of $F_t^{P,(1)}$ and $F_t^{E,(1)}$ as follows:

(i) The pair

$$(Y_t^{P,(1)}, y_t^P) = \begin{cases} \left(F_t^{P,(1)}, u_t^P - F_{t-1}^{P,(2)} x_{t-1}^{(2)} \right) & \text{if } F_t^{P,(1)} \text{ has rank } m < n_1, \\ (I_n, \widehat{x}_t^{P,(1)}) & \text{if } F_t^{P,(1)} \text{ has rank } n_1 \leq m. \end{cases}$$

(ii) The pair

$$(Y_t^{E,(1)}, y_t^E) = \begin{cases} \left(F_t^{E,(1)}, u_t^E - F_{t-1}^{E,(2)} x_{t-1}^{(2)} \right) & \text{if } F_t^{E,(1)} \text{ has rank } k < n_1, \\ (I_n, \widehat{x}_t^{E,(1)}) & \text{if } F_t^{E,(1)} \text{ has rank } n_1 \leq k. \end{cases}$$

Finally, the initial conditions are $\widehat{\Sigma}_0^i = W_0^i$, $\widetilde{\Sigma}_0^{(i,j)} = 0$, and $\widehat{\Sigma}_0^{(i,j)} = \widehat{\Sigma}_0^i + \widehat{\Sigma}_0^j$.

Theorem 4.3 (Nash Equilibrium). *Suppose Assumption 4.1 holds and there exists a unique solution $\{F_t^{P*}\}_{t=0}^{T-1}$ and $\{F_t^{E*}\}_{t=0}^{T-1}$ to (3.7)-(3.8) with $F_t^{P*,(1)}$ of rank $\min(m, n_1)$ and $F_t^{E*,(1)}$ of rank $\min(k, n_1)$. Further assume that both players apply linear policies. Then the unique Nash equilibrium policy is*

$$u_t^{i*} = F_t^{i*} y_t^i, \quad \text{with} \quad y_t^i = (\widehat{x}_t^{i,(1)}, x_t^{(2)})^\top, \quad i = P, E. \quad (4.8)$$

The corresponding optimal value functions are quadratic ($0 \leq t \leq T$):

$$V_t^P(y_t^P; F^{E*}) = (y_t^P)^\top U_t^{P*} y_t^P + \widetilde{c}_t^{P*}, \quad V_t^E(y_t^E; F^{P*}) = (y_t^E)^\top U_t^{E*} y_t^E + \widetilde{c}_t^{E*}, \quad (4.9)$$

with matrices $U_t^{P*}, U_t^{E*} \in \mathbb{R}^{n \times n}$ given in (3.9) and (3.10). Here for $i, j = P, E$ and $j \neq i$, the scalar $\widetilde{c}_t^{i*} \in \mathbb{R}$ is given by

$$\begin{aligned} \widetilde{c}_t^{i*} &= \widetilde{c}_{t+1}^{i*} + \text{Tr}(Q_t^{i,(1,1)} \widehat{\Sigma}_t^i) + \text{Tr}((\bar{L}_t^{i,(1)})^\top U_{t+1}^i \bar{L}_t^{i,(1)} \widehat{\Sigma}_t^i) + \text{Tr}((\bar{L}_t^{i,(2)})^\top U_{t+1}^i \bar{L}_t^{i,(2)} \widehat{\Sigma}_t^j) \\ &\quad + 2 \text{Tr}((\bar{L}_t^{i,(2)})^\top U_{t+1}^i \bar{L}_t^{i,(1)} \widetilde{\Sigma}_t^{(i,j)}) + \text{Tr}\left((K_{t+1}^i)^\top U_{t+1}^i K_{t+1}^i G^i\right) \\ &\quad + \text{Tr}\left(\left[(K_{t+1}^i H_{t+1}^i \Gamma_t^{(1)})^\top \quad (\Gamma_t^{(2)})^\top\right] U_{t+1}^i \begin{bmatrix} K_{t+1}^i H_{t+1}^i \Gamma_t^{(1)} \\ \Gamma_t^{(2)} \end{bmatrix} W\right), \end{aligned}$$

where $\bar{L}_t^{i,(1)} = (\widetilde{L}_t^{i,(1)}, -(A_t^{(2,1)} + B_t^{j,(2)} F_t^{j*,(1)}))^\top$, and $\bar{L}_t^{i,(2)} = (\widetilde{L}_t^{i,(2)}, B_t^{j,(2)} F_t^{j*,(1)})^\top$ with

$$\begin{aligned} \widetilde{L}_t^{i,(1)} &= -A_t^{(1,1)} \Pi_t^i - B_t^{j,(1)} F_t^{j*,(1)} - K_{t+1}^i H_{t+1}^i A_t^{(1,1)} (I - \Pi_t^i), \\ \widetilde{L}_t^{i,(2)} &= A_t^{(1,1)} \Pi_t^i + B_t^{j,(1)} F_t^{j*,(1)} - K_{t+1}^i H_{t+1}^i A_t^{(1,1)} \Pi_t^i, \end{aligned}$$

where $\Pi_t^i := (\widehat{\Sigma}_t^i - \widetilde{\Sigma}_t^{(i,j)}) (\widehat{\Sigma}_t^{(i,j)})^{-1}$. The terminal conditions are $\widetilde{c}_T^{i*} = \text{Tr}(Q_T^{i,(1,1)} \widehat{\Sigma}_T^i)$ for $i = P, E$.

5 Numerical Experiment: the Bargaining Game

In this section, we perform some numerical experiments on a bargaining game example which can be cast into the framework introduced in Section 4. Consider a two-player bargaining or negotiation game where a buyer and a seller must agree on the value of a good. Each party has a target price, that is the price they want to achieve by agreement at the deadline. The target price depends on their view of the project's true value. The buyer (resp. seller) does not know the target price of the seller (resp. buyer) or the true value of the good. The challenge is to establish a model for the bargaining situation and find the optimal bidding strategies when both parties have partial information about their counterparties, under some uncertainties (e.g. market fluctuations).

In this section we focus on the case $n_1 = 1$ when the opponent's state estimate can be inferred. The case where $n_1 = 2$, where this is not the case, has similar results and these are deferred to Appendix B.

5.1 Mathematical Set-up

We now cast this bargaining game into the mathematical framework introduced in Section 4. Assume we have two players, a buyer B and a seller S , who aim to reach an agreement on the value (or the price) of a good. The negotiation takes place over a finite period of time T . At each timestamp t , the buyer and seller simultaneously offer prices. We let $x_t^B \in \mathbb{R}$ be the price offered by the buyer and $x_t^S \in \mathbb{R}$ be the price offered by the seller. The dynamics of the offers follow

$$x_{t+1}^B = x_t^B + u_t^B + \epsilon_t^B, \quad x_{t+1}^S = x_t^S + u_t^S + \epsilon_t^S, \quad \text{with initial values } x_0^B, x_0^S, \quad (5.1)$$

Here $u_t^B \in \mathbb{R}$ is the change in the buyer's offer and $u_t^S \in \mathbb{R}$ is the change in the seller's offer at time t . The random variables ϵ_t^B and ϵ_t^S are IID, representing the noise in both parties offers with $\epsilon_t^B \sim \mathcal{N}(0, \overline{W}^B)$ and $\epsilon_t^S \sim \mathcal{N}(0, \overline{W}^S)$, respectively. We note that ϵ_t^B and ϵ_t^S serve as regularization terms to guarantee the non-degeneracy of the state noise. Another way of thinking about this is to consider u_t^B and u_t^S as the *intended* change of their offers when players can not completely control the differences between their offers (for example, due to some external restrictions). Both players can observe each other's exact offers. Thus $(x_t^B, x_t^S)^\top$ corresponds to the fully observable part $x_t^{(2)}$ in Section 4.

We assume the value of the good $p_t \in \mathbb{R}$ is not available to both players and its dynamics follow:

$$p_{t+1} = p_t + w_t, \quad (5.2)$$

where $\{w_t\}_{t=0}^{T-1}$ is a sequence of IID Gaussian random variables with zero mean and covariance $\overline{W} \in \mathbb{R}$. Both the buyer and the seller do not have access to the true value of the good. Instead, they observe a noisy version of the value using their private information. At time $t = 0$, player i ($i = B, S$) believes that the initial value $p_0 \sim \mathcal{N}(\hat{p}_0^i, W_0^i)$, and after that player i observes the following noisy signal:

$$z_{t+1}^i = p_{t+1} + w_{t+1}^i, \quad w_{t+1}^i \sim \mathcal{N}(0, G^i), \quad t = 0, 1, \dots, T-1, \quad (5.3)$$

where $\{w_t^i\}_{t=1}^{T-1}$ is a sequence of IID random variables, and $\{w_t^B\}_{t=1}^{T-1}$ and $\{w_t^S\}_{t=1}^{T-1}$ are independent of each other. Thus p_t corresponds to the partially observable part $x_t^{(1)}$ in (4.1) of Section 4, with $n_1 = 1$.

We formulate player i 's ($i = B, S$) objective in the game as

$$\min_{\{u_t^i\}_{t=0}^{T-1}} \mathbb{E} \left[\alpha_i (x_T^B - x_T^S)^2 + \beta_i (x_T^i - (1 + \delta_i)p_T)^2 + \sum_{t=0}^{T-1} R_t^i (u_t^i)^2 \middle| \mathcal{H}_0^i \right], \quad (5.4)$$

where $\delta_B \in (-1, 0)$ and $\delta_S \in (0, 1)$ are the scalars that determine the buyer's and the seller's target price at terminal time T . The constants $\alpha_B > 0$ and $\alpha_S > 0$ are the penalties for not reaching an agreement, and $\beta_B > 0$ and $\beta_S > 0$ are the penalties for deviating from their target prices. The quadratic terms $\alpha_S (x_T^B - x_T^S)^2$ and $\alpha_B (x_T^B - x_T^S)^2$ can be viewed as a relaxation of the hard constraint $x_T^B = x_T^S$. The parameters $R_t^B > 0$ and $R_t^S > 0$ measure the cost of adjusting the offer price at each time step, thus the final terms represent the penalty for making concessions. The filtrations $\mathcal{H}_0^B := \{\hat{\xi}_0^B, W_0^B, W_0^S\}$ and $\mathcal{H}_0^S := \{\hat{\xi}_0^S, W_0^B, W_0^S\}$ represent the information available at time 0.

Both players have the incentive to reach an agreement at terminal time T . The desire to reach this agreement is characterized by the value of α_B and α_S , which may be different for the buyer and the seller. The hard constraint $x_T^B = x_T^S$ can be recovered by letting α_B and α_S tend to infinity. The seller wants to sell the good at a price that is higher than (his estimate of) the true price and thus $\delta_S > 0$. Similarly $\delta_B < 0$ as the buyer has the incentive to buy at a price lower than his estimated true price.

5.2 Experiments

In this section, we present some numerical experiments and discuss the effect of observation noise and our information corrections for the bargaining game introduced in Section 5.1. We focus on the case

$n_1 = 1$, where the dynamics of the value of the good and the players' noisy observations are defined in (5.2)-(5.3). The bargaining model considered in this section satisfies the conditions for the special case described in point 5. of Remark 2.5, where each player can fully recover the opponent's state estimate in the previous step.

Experimental Set-up. In the bargaining game (5.1)-(5.4), the model parameters are,

$$A_t = I, \quad B_t^B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad B_t^S = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad W = \begin{bmatrix} \bar{W} & 0 & 0 \\ 0 & \bar{W}^B & 0 \\ 0 & 0 & \bar{W}^S \end{bmatrix}, \quad Q_t^B = Q_t^S = 0, \text{ and}$$

$$Q_T^B = \begin{bmatrix} \beta_B(1 + \delta_B)^2 & -\beta_B(1 + \delta_B) & 0 \\ -\beta_B(1 + \delta_B) & \alpha_B + \beta_B & -\alpha_B \\ 0 & -\alpha_B & \alpha_B \end{bmatrix}, \quad Q_T^S = \begin{bmatrix} \beta_S(1 + \delta_S)^2 & 0 & -\beta_S(1 + \delta_S) \\ 0 & \alpha_S & -\alpha_S \\ -\beta_S(1 + \delta_S) & -\alpha_S & \alpha_S + \beta_S \end{bmatrix},$$

for $t = 0, 1, \dots, T - 1$. Also we have $H_t^i = I$ for $i = S, B$.

In the experiments we let $\alpha_B = \alpha_S = 50$, $\beta_B = \beta_S = 30$, $\delta_B = -0.05$, $\delta_S = 0.05$, and $T = 10$, so the players care more about reaching an agreement with each other. We set the penalty function to be $R_t^i = \rho_i \exp(-\gamma_i t)$ for $i = B, S$ with $\rho_B = \rho_S = 15$ and $\gamma_B = \gamma_S = 0.1$. The penalty function decays over time which allows players to be more flexible near the deadline to reach an agreement. For the initial state we set $p_0 = 50$, $x_0^B = 10$, $x_0^S = 90$. We also set $\bar{W} = 9$ for the noise in the dynamics of the true value of the good, and $\bar{W}^B = \bar{W}^S = 10^{-12}$. The reason for adding the small noise to x_t^B and x_t^S is to guarantee the well-definedness of the problem. In practice we can set $\bar{W}^B = \bar{W}^S = 0$, and the numerical experiments will still work.

To see the effect of the observation noise, we let the buyer have a much more noisy observation of the true price ($G^B = 100$ and $G^S = 1$). We also set $\hat{x}_0^B = 40$ with $W_0^B = 100$ for the buyer and $\hat{x}_0^S = 51$ with $W_0^S = 1$ for the seller, thus the buyer has a far more inaccurate guess of the initial state. In the figures and tables we will write IC for information corrections.

Effect of Observation Noise. Since the buyer receives relatively noisy signals of the true price, their price estimate (indicated in orange) will be more inaccurate than the seller's (indicated in blue) in the example shown in Figure 1. The behaviour of both players is similar to that in the full information case, since the buyer utilizes the seller's accurate information to improve their own state estimate.

Effect of Information Corrections. A key contribution of our work is information corrections, where players correct their estimate of the state after observing their opponent's actions. We demonstrate the power of the information corrections in Figure 2. When the buyer skips steps (4.7a)-(4.7c), their state estimate will rely purely on their own observations and thus can be very inaccurate. However, with information corrections, they can obtain a better estimate which is less affected by the noisy observations. Hence they are more likely to reach an agreement with the seller at a reasonable price.

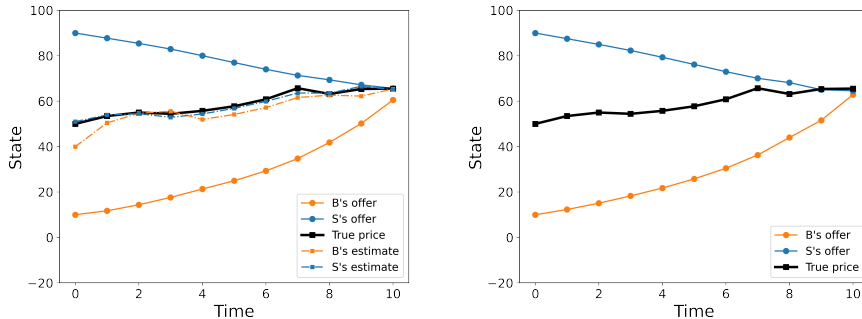


Figure 1: Comparison between the full observation (right) and partial observation (left) cases.

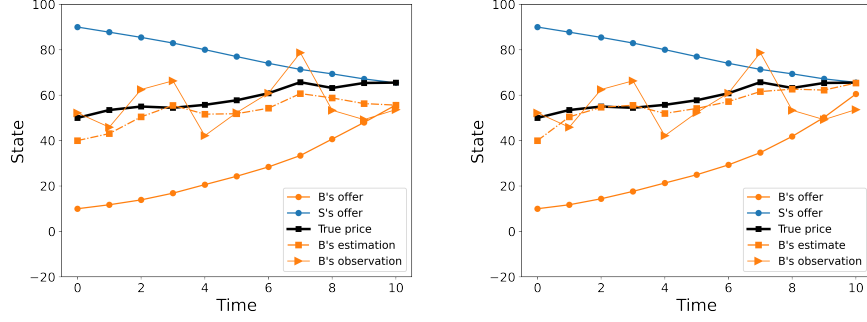


Figure 2: The buyer’s price estimate with IC (right) and without IC (left).

We now show some statistics for the buyer’s estimation error in Table 1, where both the average mean squared error and average mean absolute error of 500 experiments (each experiment consists of 10 rounds of bargaining) are smaller when information corrections are used. We also show the effect of using information corrections on the outcomes of the bargaining game in Table 2. If the difference between the players’ final offers is less than 3, we consider them to have reached an agreement. We can see that with information corrections the players are more likely to reach an agreement. The difference is shown in Figure 2. We can see that with information correction the final offers of the buyer and the seller are closer to each other and they have made the deal, while without information correction they are not able to reach an agreement in the end.

	Mean squared error	Mean absolute error
With information correction	17.43	3.10
Without information correction	35.91	4.83

Table 1: Effect of IC on the buyer’s estimation error (average over 500 experiments).

Furthermore, the above setting is an “asymmetric” case, where the buyer has a more inaccurate estimate of the initial state and receives noisier signals during the negotiation. We now compare the number of agreements obtained in this case (Table 2) with that obtained in two symmetric cases, which can be considered as benchmarks. In the symmetric case where both players have “accurate” information with small observation noises, we let $G^B = G^S = 1$ for both players, and set $\hat{x}_0^B = 49$, and $\hat{x}_0^S = 51$ with $W_0^B = W_0^S = 1$; in the symmetric where they both have inaccurate information, we set $G^B = G^S = 100$, $\hat{x}_0^B = 40$, and $\hat{x}_0^S = 60$ with $W_0^B = W_0^S = 100$. We run 500 experiments in each case, and the number of agreements when they both have “accurate” information is the same regardless of whether they utilize the observed actions from their opponent. However, when both players receive very noisy signals, having information corrections significantly improves the number of agreements. This further demonstrates the need for incorporating information corrections especially when players have different levels of observation noise, as is often the case in practice since players may have a variety of different information sources. We also note that the number of agreements reached in the asymmetric case is closer to that in the inaccurate symmetric case. Although the players have improved their state estimate at the previous time step, when the players make offers they are based on their current noisy observation.

	Asymmetric	Symmetric (“accurate”)	Symmetric (inaccurate)
With IC	371	470	367
Without IC	253	470	247

Table 2: Number of agreements achieved in 500 experiments with and without IC.

We can also compare the players’ cost in the asymmetric and symmetric cases (see Table 3). The

costs are calculated empirically based on (5.4) In the asymmetric case and the symmetric case when they both have very noisy observations, both players achieve significantly lower costs when using information corrections.

	Asymmetric	Symmetric (“accurate”)	Symmetric (inaccurate)
With IC (Buyer/Seller)	2250/2235	1923/2053	2505/2715
Without IC (Buyer/Seller)	3068/2819	1924/2053	3244/3359

Table 3: Players’ average costs with and without IC in 500 experiments.

Aiming at Beneficial Prices. In the above experiments we focused on the case where both players strive to reach an agreement with their opponent ($\alpha_i > \beta_i$ for $i = B, S$). However, in some situations players may be more keen on achieving their target price or a more beneficial price in negotiations and not be so concerned if an agreement is reached. We now let the buyer focus more on pursuing their target price and let the seller mainly seek an agreement with the buyer. We let $\alpha_B = 20$, $\alpha_S = 50$, $\beta_B = 40$, $\beta_S = 30$, $\rho_B = \rho_S = 10$, and $\overline{W} = 1$. We also let the buyer have far more inaccurate information than the seller by setting $G^B = 100$, $G^S = 1$, $\hat{x}_0^B = 70$, $W_0^B = 100$, $\hat{x}_0^S = 53$, and $W_0^S = 1$. Other parameters are set to be the same as in the previous experiments.

In Table 4, we can see that the information corrections significantly improve the number of agreements achieved. Here we set the agreement price achieved to be the average of x_T^B and x_T^S . For a fair comparison we only consider situations where an agreement is achieved in both cases (with and without information corrections). We observe that there is a gap between the confidence intervals of the agreement prices, which illustrates that the buyer can obtain a better deal by using the information corrections to more effectively exploit the seller’s willingness to sacrifice their target price in their desire to reach an agreement.

	Number of agreements	Mean of APs	95% confidence interval of APs
With IC	442	48.58	(48.31, 48.84)
Without IC	342	49.61	(49.34, 49.88)

Table 4: Bargaining outcomes and agreement prices (APs) with and without IC in 500 experiments.

References

- [1] Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica: Journal of the Econometric Society*, pages 1041–1063, 1990.
- [2] Tamer Başar and Geert Jan Olsder. *Dynamic Non-Cooperative Game Theory*. Society for Industrial and Applied Mathematics, 1998.
- [3] Ulrich Doraszelski and Juan F Escobar. A theory of regular Markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification. *Theoretical Economics*, 5(3):369–402, 2010.
- [4] Qi Fu, Chee-Khian Sim, and Chung-Piaw Teo. Profit sharing agreements in decentralized supply chains: A distributionally robust approach. *Operations Research*, 66(2):500–513, 2018.
- [5] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT press, 1991.
- [6] Abhishek Gupta, Cédric Langbort, and Tamer Başar. Dynamic games with asymmetric information and resource constrained players with applications to security of cyberphysical systems. *IEEE Transactions on Control of Network Systems*, 4(1):71–81, 2016.
- [7] Abhishek Gupta, Ashutosh Nayyar, Cédric Langbort, and Tamer Basar. Common information based Markov perfect equilibria for linear-Gaussian games with asymmetric information. *SIAM Journal on Control and Optimization*, 52(5):3228–3260, 2014.
- [8] Nasimeh Heydaribeni and Achilleas Anastasopoulos. Structured equilibria for dynamic games with asymmetric information and dependent types. *arXiv preprint arXiv:2009.04253*, 2020.
- [9] Mingming Leng and Mahmut Parlar. Allocation of cost savings in a three-level supply chain with demand information sharing: A cooperative-game approach. *Operations Research*, 57(1):200–213, 2009.
- [10] Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219, 2001.
- [11] Ashutosh Nayyar, Abhishek Gupta, Cedric Langbort, and Tamer Başar. Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games. *IEEE Transactions on Automatic Control*, 59(3):555–570, 2013.
- [12] Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic games with asymmetric information: Common information based perfect Bayesian equilibria and sequential decomposition. *IEEE Transactions on Automatic Control*, 62(1):222–237, 2016.
- [13] Meir Pachter. LQG dynamic games with a control-sharing information pattern. *Dynamic Games and Applications*, 7(2):289–322, 2017.
- [14] Meir Pachter and Khanh Pham. Informational issues in decentralized control. In *Dynamics of information systems: computational and mathematical challenges*, pages 45–76. Springer, 2014.
- [15] Ariel Pakes and Paul McGuire. Stochastic algorithms, symmetric Markov perfect equilibrium, and the “curse” of dimensionality. *Econometrica*, 69(5):1261–1281, 2001.
- [16] Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, pages 97–109, 1982.

- [17] Robert F Stengel. *Optimal control and estimation*. Courier Corporation, 1994.
- [18] Deepanshu Vasal and Achilleas Anastasopoulos. Signaling equilibria for dynamic LQG games with asymmetric information. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 6901–6908. IEEE, 2016.
- [19] Deepanshu Vasal, Abhinav Sinha, and Achilleas Anastasopoulos. A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information. *IEEE Transactions on Automatic Control*, 64(1):81–96, 2018.
- [20] Hans S Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147, 1968.
- [21] Xianrong Zheng, Patrick Martin, Wendy Powley, and Kathryn Brohman. Applying bargaining game theory to web services negotiation. In *2010 IEEE International Conference on Services Computing*, pages 218–225. IEEE, 2010.

A Tower Property

We return to the calculation of (2.65) and (2.66). Recall that we consider the case when F_t^P has rank $m < n$ and F_t^E has rank $k < n$ for all $t = 0, 1, \dots, T-1$.

Matching LHS with RHS. We can see the first term (quadratic in \widehat{x}_{T-1}^P) on the LHS and RHS are the same. Thus we are left to show that the constants terms in (2.65) and (2.66) are the same. We prove this in the following steps.

Step 1: Expanding L_1 and L_2 on the LHS. We first calculate the three terms in (2.65) that involves L_1 and L_2 using (2.63) and (2.64). The first term is given by

$$\begin{aligned} \text{Tr} \left(L_1^\top Q_T^P L_1 \widehat{\Sigma}_{T-1}^E \right) &= \underbrace{\text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P B_{T-1}^E F_{T-1}^E \widehat{\Sigma}_{T-1}^E \right)}_{(1a)} \\ &\quad + \underbrace{\text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \widehat{\Sigma}_{T-1}^E \right)}_{(1b)} \\ &\quad + 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E \widehat{\Sigma}_{T-1}^E \right) \end{aligned}$$

For the second term, we first note that by adding and subtracting A_{T-1} , we obtain

$$\begin{aligned} L_2 &= -(I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P - B_{T-1}^E F_{T-1}^E - K_T^P H_T^P A_{T-1} \\ &= -(I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P - B_{T-1}^E F_{T-1}^E + (I - K_T^P H_T^P) A_{T-1} - A_{T-1} \end{aligned}$$

Then, the second term is given by

$$\begin{aligned} \text{Tr} \left(L_2^\top Q_T^P L_2 \widehat{\Sigma}_{T-1}^P \right) &= \underbrace{\text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \widehat{\Sigma}_{T-1}^P \right)}_{(2a)} \\ &\quad + \underbrace{\text{Tr} \left((A_{T-1} + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \widehat{\Sigma}_{T-1}^P \right)}_{(2b)} \\ &\quad + \underbrace{\text{Tr} \left(((I - K_T^P H_T^P) A_{T-1})^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widehat{\Sigma}_{T-1}^P \right)}_{(2c)} \\ &\quad + 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \widehat{\Sigma}_{T-1}^P \right) \\ &\quad - \underbrace{2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widehat{\Sigma}_{T-1}^P \right)}_{(2d)} \\ &\quad - 2 \text{Tr} \left((A_{T-1} + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widehat{\Sigma}_{T-1}^P \right) \end{aligned}$$

Similarly the third term is given by

$$\begin{aligned} 2 \text{Tr} \left(L_1^\top Q_T^P L_2 \widetilde{\Sigma}_{T-1}^{(P,E)} \right) &= \underbrace{-2 \text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \widetilde{\Sigma}_{T-1}^{(P,E)} \right)}_{(3a)} \\ &\quad - 2 \text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \widetilde{\Sigma}_{T-1}^{(P,E)} \right) \\ &\quad + 2 \text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widetilde{\Sigma}_{T-1}^{(P,E)} \right) \\ &\quad - 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \widetilde{\Sigma}_{T-1}^{(P,E)} \right) \end{aligned}$$

$$\begin{aligned}
& - \underbrace{2 \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \widetilde{\Sigma}_{T-1}^{(P,E)} \right)}_{(3b)} \\
& + \underbrace{2 \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widetilde{\Sigma}_{T-1}^{(P,E)} \right)}_{(3c)}
\end{aligned}$$

Now we observe that all the constant terms on RHS except $\operatorname{Tr}(\Gamma_{T-1}^\top Q_T \Gamma_{T-1} W)$ can be cancelled with (2b), (1a), and (3a) on LHS, that is

$$\text{Constant terms on RHS} = (2b) + (1a) + (3a) + \operatorname{Tr}(\Gamma_{T-1}^\top Q_T \Gamma_{T-1} W). \quad (\text{A.1})$$

Therefore, our goal now is to show that $\operatorname{Tr}(\Gamma_{T-1}^\top Q_T \Gamma_{T-1} W)$ equals the rest of terms on LHS. Before matching the terms, we first merge and simplify some of the terms on LHS.

Step 2: Merging quadratic terms of Π_{T-1}^P on LHS. Recall that Π_{T-1}^P is defined as $\Pi_{T-1}^P = (\widehat{\Sigma}_{t-1}^P - \widetilde{\Sigma}_{t-1}^{(P,E)}) (\widehat{\Sigma}_{t-1}^{(P,E)})^{-1}$. Thus we have

$$\Pi_{T-1}^P \widehat{\Sigma}_{t-1}^{(P,E)} = \widehat{\Sigma}_{t-1}^P - \widetilde{\Sigma}_{t-1}^{(P,E)}. \quad (\text{A.2})$$

This provides a way to reduce the order of Π_{T-1}^P on LHS. Collecting terms which are quadratic in Π_{T-1}^P we obtain

$$\begin{aligned}
(1b) + (2a) + (3b) &= \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \cdot \right. \\
&\quad \left. (\widehat{\Sigma}_{T-1}^P + \widehat{\Sigma}_{T-1}^E - \widetilde{\Sigma}_{T-1}^{(P,E)} - \widetilde{\Sigma}_{T-1}^{(E,P)}) \right) \\
&= \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \widehat{\Sigma}_{T-1}^{(P,E)} \right), \quad (\text{A.3}) \\
&= \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\widehat{\Sigma}_{t-1}^P - \widetilde{\Sigma}_{t-1}^{(P,E)}) \right) \quad (\text{A.4})
\end{aligned}$$

where (A.3) holds by definition of $\widehat{\Sigma}_{T-1}^{(P,E)}$ given in (2.38), and (A.4) holds by (A.2). Now we observe that the term in (A.4) will be cancelled with a half of the sum (2d) + (3c), since

$$(2d) + (3c) = 2 \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\widetilde{\Sigma}_{T-1}^{(P,E)} - \widehat{\Sigma}_{T-1}^P) \right).$$

Therefore in summary, we have

$$(1b) + (2a) + (3b) + (2d) + (3c) = \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\widetilde{\Sigma}_{T-1}^{(P,E)} - \widehat{\Sigma}_{T-1}^P) \right).$$

Step 3: Merging the last three terms on LHS. Going back to (2.65), we now consider the last three constant terms that do not involve L_1 and L_2 . We now add the first two terms to (2c), which leads to

$$\begin{aligned}
& \operatorname{Tr}((K_T^P)^\top Q_T^P K_T^P G^P) + \operatorname{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) + (2c) \\
&= \operatorname{Tr} \left((K_T^P)^\top Q_T^P K_T^P (H_T^P \Gamma_{T-1} W \Gamma_{T-1}^\top (H_T^P)^\top + G^P) \right) \\
&\quad + \operatorname{Tr} \left(((I - K_T^P H_T^P) A_{T-1})^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \widehat{\Sigma}_{T-1}^P \right) \\
&= \operatorname{Tr} \left((K_T^P)^\top Q_T^P K_T^P (H_T^P \Gamma_{T-1} W \Gamma_{T-1}^\top (H_T^P)^\top + G^P + H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P A_{T-1}^\top (H_T^P)^\top) \right) \\
&\quad - 2 \operatorname{Tr}(A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) + \operatorname{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) \quad (\text{A.5})
\end{aligned}$$

To simplify the first term in the above sum, we use the following two facts.

Fact 1. We first note that

$$(\widehat{\Sigma}_T^P)^- = A_{T-1}(\widehat{\Sigma}_{T-1}^P)^+ A_{T-1}^\top + \Gamma_{T-1} W \Gamma_{T-1}^\top \quad (\text{A.6})$$

$$= A_{T-1} \left(\widehat{\Sigma}_{T-1}^P - (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\widehat{\Sigma}_{T-1}^{(P,E)})^{-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)})^\top \right) A_{T-1}^\top + \Gamma_{T-1} W \Gamma_{T-1}^\top \quad (\text{A.7})$$

where (A.6) holds by (2.21e) and (A.7) holds by (2.21c). Thus by rearranging terms in (A.6) and since $\Pi_{T-1}^P = (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\widehat{\Sigma}_{T-1}^{(P,E)})^{-1}$, we have

$$A_{T-1} \widehat{\Sigma}_{T-1}^P A_{T-1}^\top + \Gamma_{T-1} W \Gamma_{T-1}^\top = (\widehat{\Sigma}_T^P)^- + A_{T-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top. \quad (\text{A.8})$$

Fact 2. By definition of K_t^i in (2.21f), we have

$$K_T^P \left(H_T^P (\widehat{\Sigma}_T^P)^- (H_T^P)^\top + G^P \right) = (\widehat{\Sigma}_T^P)^- (H_T^P)^\top.$$

Therefore,

$$(K_T^P)^\top Q_T^P K_T^P \left(H_T^P (\widehat{\Sigma}_T^P)^- (H_T^P)^\top + G^P \right) = (K_T^P)^\top Q_T^P (\widehat{\Sigma}_T^P)^- (H_T^P)^\top. \quad (\text{A.9})$$

Having the above two facts, we can plug in (A.8) and (A.9) into (A.5) to get

$$\begin{aligned} & \text{Tr}((K_T^P)^\top Q_T^P K_T^P G^P) + \text{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) + (2c) \\ = & \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P \left(H_T^P ((\widehat{\Sigma}_T^i)^- + A_{T-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top) (H_T^P)^\top + G^P \right) \right) \\ & - 2 \text{Tr}(A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) + \text{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) \\ = & \text{Tr} \left((K_T^P)^\top Q_T^P (\widehat{\Sigma}_T^P)^- (H_T^P)^\top \right) + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top (H_T^P)^\top \right) \\ & - 2 \text{Tr}(A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) + \text{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) \\ = & \text{Tr} \left(Q_T^P ((\widehat{\Sigma}_T^P)^- - \widehat{\Sigma}_T^P) \right) + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top (H_T^P)^\top \right) \\ & - 2 \text{Tr}(A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) + \text{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) \end{aligned} \quad (\text{A.10})$$

where the last equation holds since $(\widehat{\Sigma}_T^P)^- (H_T^P)^\top (K_T^P)^\top = (\widehat{\Sigma}_T^P)^- - \widehat{\Sigma}_T^P$ by (2.21h). Now by (A.10), the sum of (2c) and the last three terms in (2.65) is given by

$$\begin{aligned} & \text{Tr}((K_T^P)^\top Q_T^P K_T^P G^P) + \text{Tr}(\Gamma_{T-1}^\top (H_T^P)^\top (K_T^P)^\top Q_T^P K_T^P H_T^P \Gamma_{T-1} W) + \text{Tr}(Q_T^P \widehat{\Sigma}_T^P) + (2c) \\ = & \text{Tr} \left(Q_T^P (\widehat{\Sigma}_T^P)^- \right) + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\widehat{\Sigma}_{T-1}^P - \widetilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top (H_T^P)^\top \right) \\ & - 2 \text{Tr}(A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) + \text{Tr}(A_{T-1}^\top Q_T^P A_{T-1} \widehat{\Sigma}_{T-1}^P) \end{aligned} \quad (\text{A.11})$$

Step 4: Collecting all the terms on LHS. Recall that (2b) + (1a) + (3a) can be cancelled with RHS in Step 1, the sum of (1b) + (2a) + (3b) + (2d) + (3c) is given in Step 2, and the sum of the last three terms on LHS and (2c) is given in Step 3. We are now ready to sum up all the terms on LHS using the simplified forms obtained from Steps 2 and 3.

$$\begin{aligned} & \text{Constant terms on LHS} \\ = & (2b) + (1a) + (3a) \\ & + \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\widetilde{\Sigma}_{T-1}^{(P,E)} - \widehat{\Sigma}_{T-1}^P) \right) \\ & + 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E \widehat{\Sigma}_{T-1}^E \right) \end{aligned} \quad (\text{A.12})$$

$$+2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \hat{\Sigma}_{T-1}^P \right) \quad (\text{A.13})$$

$$-2 \text{Tr} \left((A_{T-1} + B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \hat{\Sigma}_{T-1}^P \right) \quad (\text{A.14})$$

$$-2 \text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P \tilde{\Sigma}_{T-1}^{(P,E)} \right) \quad (\text{A.15})$$

$$+2 \text{Tr} \left((B_{T-1}^E F_{T-1}^E)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \tilde{\Sigma}_{T-1}^{(P,E)} \right) \quad (\text{A.16})$$

$$-2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (A_{T-1} + B_{T-1}^E F_{T-1}^E) \tilde{\Sigma}_{T-1}^{(P,E)} \right) \quad (\text{A.17})$$

$$+ \text{Tr} \left(Q_T^P (\hat{\Sigma}_T^P)^- \right) + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top (H_T^P)^\top \right) \\ -2 \text{Tr} (A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P).$$

We observe that by adding (part of the terms in) (A.12), (A.13), (A.15), and (A.17) together, we can obtain the sum

$$2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E (\hat{\Sigma}_{T-1}^P + \hat{\Sigma}_{T-1}^E - \tilde{\Sigma}_{T-1}^{(P,E)} - \tilde{\Sigma}_{T-1}^{(E,P)}) \right) \\ = 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P B_{T-1}^E F_{T-1}^E \hat{\Sigma}_{T-1}^{(P,E)} \right) \\ = 2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1})^\top Q_T^P B_{T-1}^E F_{T-1}^E (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(E,P)}) \right)$$

where in the last equation we use the trick (A.2) again. This sum will be further cancelled with part of (A.14) and (A.16). After these manipulations we have

Constant terms on LHS

$$= (2b) + (1a) + (3a) \\ + \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\tilde{\Sigma}_{T-1}^{(P,E)} - \hat{\Sigma}_{T-1}^P) \right) \quad (\text{A.18})$$

$$+2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P \right) \quad (\text{A.19})$$

$$-2 \text{Tr} \left(A_{T-1}^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \hat{\Sigma}_{T-1}^P \right) \quad (\text{A.20})$$

$$-2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} \tilde{\Sigma}_{T-1}^{(P,E)} \right) \quad (\text{A.21})$$

$$+ \text{Tr} \left(Q_T^P (\hat{\Sigma}_T^P)^- \right) + \text{Tr} \left((K_T^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top A_{T-1}^\top (H_T^P)^\top \right) \\ -2 \text{Tr} (A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P). \quad (\text{A.22})$$

Now adding terms in (A.20) and (A.22) together we have

$$-2 \text{Tr} \left(A_{T-1}^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} \hat{\Sigma}_{T-1}^P \right) - 2 \text{Tr} (A_{T-1}^\top Q_T^P K_T^P H_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) \\ + \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) \\ = - \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P)$$

Also, adding (A.18), (A.19), and (A.21) together leads to

$$\text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P (I - K_T^P H_T^P) A_{T-1} (\tilde{\Sigma}_{T-1}^{(P,E)} - \hat{\Sigma}_{T-1}^P) \right) \\ +2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P \right) \\ -2 \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} \tilde{\Sigma}_{T-1}^{(P,E)} \right) \\ = - \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\tilde{\Sigma}_{T-1}^{(P,E)} - \hat{\Sigma}_{T-1}^P) \right) \\ + \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) \right)$$

Therefore, the constant terms on LHS can be further simplified as

$$\begin{aligned}
& \text{Constant terms on LHS} \\
&= (2b) + (1a) + (3a) \\
&\quad - \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\tilde{\Sigma}_{T-1}^{(P,E)} - \hat{\Sigma}_{T-1}^P) \right) \\
&\quad + \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) \right) \\
&\quad - \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} \left(Q_T^P (\hat{\Sigma}_T^P)^\top \right) \\
&\quad + \text{Tr} \left((K_T^P H_T^P A_{T-1} \Pi_{T-1}^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) \right) \\
&= (2b) + (1a) + (3a) \\
&\quad - \text{Tr} \left((A_{T-1} \Pi_{T-1}^P)^\top Q_T^P K_T^P H_T^P A_{T-1} (\tilde{\Sigma}_{T-1}^{(P,E)} - \hat{\Sigma}_{T-1}^P) \right) \\
&\quad + \text{Tr} \left(((I - K_T^P H_T^P) A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) \right) \\
&\quad - \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} \left(Q_T^P (\hat{\Sigma}_T^P)^\top \right) \\
&= (2b) + (1a) + (3a) \\
&\quad + \text{Tr} \left((A_{T-1} \Pi_{T-1}^P)^\top Q_T^P A_{T-1} (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) \right) \\
&\quad - \text{Tr} (A_{T-1}^\top Q_T^P A_{T-1} \hat{\Sigma}_{T-1}^P) + \text{Tr} \left(Q_T^P (\hat{\Sigma}_T^P)^\top \right) \\
&= (2b) + (1a) + (3a) + \text{Tr} (\Gamma_{T-1}^\top Q_T^P \Gamma_{T-1} W) \tag{A.23}
\end{aligned}$$

where the second last equality holds since $(\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)}) (\Pi_{T-1}^P)^\top = \Pi_{T-1}^P (\hat{\Sigma}_{T-1}^P - \tilde{\Sigma}_{T-1}^{(P,E)})^\top$, and the last equality holds by (A.7).

Combining (A.23) and (A.1), we can see that the constant terms on LHS and RHS are the same. Recall that we have already matched the terms that are quadratic in \hat{x}_{T-1}^P on both sides. Therefore, the tower property given in (2.59) holds.

B Bargaining Game with $n_1 \geq 2$

In Section 5 we examined a bargaining model where the players only have noisy observations of the true value of the good ($n_1 = 1$), and they can fully recover the opponent's state estimate. In this section we introduce a more complex setting, where the true value of the good depends on a set of factors and the players can only observe noisy versions of the factors. We will first construct the bargaining model which involves high dimensional factors, and then focus on the case when $n_1 = 2$. In contrast to Section 5, here players cannot fully recover their opponent's state estimate. We will finally present some numerical results to show the effect of observation noise and information corrections.

Dynamics of the Value of the Good When $n_1 > 1$. Assume that the value of the good takes the form

$$p_t = \langle \theta, \xi_t \rangle, \tag{B.1}$$

with $\xi_t \in \mathbb{R}^{n_1}$ a set of factors that determines the value of the good with coefficient $\theta \in \mathbb{R}^{n_1}$. For the value of common commodities, the factors could include weather, government policies, international events, consumer preferences, shifting input costs, and supply and demand imbalance. We assume the factors follow a simple model:

$$\xi_{t+1} = \xi_t + w_t,$$

where $\{w_t\}_{t=0}^{T-1}$ is a sequence of IID Gaussian random variables with zero mean and covariance $\overline{W} \in \mathbb{R}^{n_1 \times n_1}$.

Both the buyer and the seller do not have access to the true value of the good nor the precise value of the factors. Instead, they observe a noisy version of the factors using their private information. At time $t = 0$, player i ($i = B, S$) believes that the initial factor signal:

$$\xi_0 \sim \mathcal{N}(\widehat{\xi}_0^i, W_0^i), \quad (\text{B.2})$$

and thereafter player i observes the following noisy factor signal:

$$z_{t+1}^i = \xi_{t+1} + w_{t+1}^i, \quad w_{t+1}^i \sim \mathcal{N}(0, G^i), \quad t = 0, 1, \dots, T-1, \quad (\text{B.3})$$

where $\{w_t^i\}_{t=1}^{T-1}$ is a sequence of IID random variables, and $\{w_t^B\}_{t=1}^{T-1}$ and $\{w_t^S\}_{t=1}^{T-1}$ are independent.

We Note that this case does not degenerate to the $n_1 = 1$ case as players have different noisy factor signals leading to different value estimates

The Bargaining Model when $n_1 = 2$. Now we focus on the case where the value of the good depends on a set of two factors and players can only observe a noisy version of the factors. The dynamics of the value of the good and the noisy signals are defined in (B.1)-(B.3). In contrast to the previous section, here the players are not able to fully recover their opponent's state estimate.

We set the coefficient $\theta = (\theta_1, \theta_2)^\top$ and the model parameters to be, for $t = 0, 1, \dots, T-1$

$$A_t = I, \quad B_t^B = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad B_t^S = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad W = \begin{bmatrix} \overline{W}^1 & 0 & 0 & 0 \\ 0 & \overline{W}^2 & 0 & 0 \\ 0 & 0 & \overline{W}^B & 0 \\ 0 & 0 & 0 & \overline{W}^S \end{bmatrix}, \quad Q_t^B = Q_t^S = 0.$$

and

$$Q_T^B = \begin{bmatrix} \beta_B(1+\delta_B)^2\theta_1^2 & \beta_B(1+\delta_B)^2\theta_1\theta_2 & -\beta_B(1+\delta_B)\theta_1 & 0 \\ \beta_B(1+\delta_B)^2\theta_1\theta_2 & -\beta_B(1+\delta_B)^2\theta_2^2 & -\beta_B(1+\delta_B)\theta_2 & 0 \\ -\beta_B(1+\delta_B)\theta_1 & -\beta_B(1+\delta_B)\theta_2 & \alpha_B + \beta_B & -\alpha_B \\ 0 & 0 & -\alpha_B & \alpha_B \end{bmatrix},$$

$$Q_T^S = \begin{bmatrix} \beta_S(1+\delta_S)^2\theta_1^2 & \beta_S(1+\delta_S)^2\theta_1\theta_2 & 0 & -\beta_S(1+\delta_S)\theta_1 \\ \beta_S(1+\delta_S)^2\theta_1\theta_2 & \beta_S(1+\delta_S)^2\theta_2^2 & 0 & -\beta_S(1+\delta_S)\theta_2 \\ 0 & 0 & \alpha_S & -\alpha_S \\ -\beta_S(1+\delta_S)\theta_1 & -\beta_S(1+\delta_S)\theta_2 & -\alpha_S & \alpha_S + \beta_S \end{bmatrix}.$$

Also we have $H_t^i = I$ for $i = S, B$.

For comparison we let $\alpha_B, \alpha_S, \beta_B, \beta_S, \delta_B, \delta_S, T$, and the penalty function to be the same as in Section. We also set $\theta_1 = \theta_2 = 1$. For the initial state we set $\xi_0 = (30, 20)^\top$, $x_0^B = 10$, $x_0^S = 90$. We also set $\overline{W}^1 = \overline{W}^2 = 4.5$ for the noise in the dynamics of the factors, and $\overline{W}^B = \overline{W}^S = 10^{-12}$.

To see the effect of the observation noise, we let the buyer have a much noisier observation of the true price by setting

$$G^B = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}, \quad G^S = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

For the players' belief about the initial state, we set $\widehat{x}_0^B = (25, 15)^\top$ and $\widehat{x}_0^S = (31, 20)^\top$ with

$$W_0^B = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}, \quad W_0^S = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},$$

thus the buyer has a far more inaccurate guess of the initial state.

Effect of Observation Noise. As for the case when $n_1 = 1$, the seller has a more accurate state estimate since he receives less noisy signals. The behaviour of the players is similar to that in the full information case.

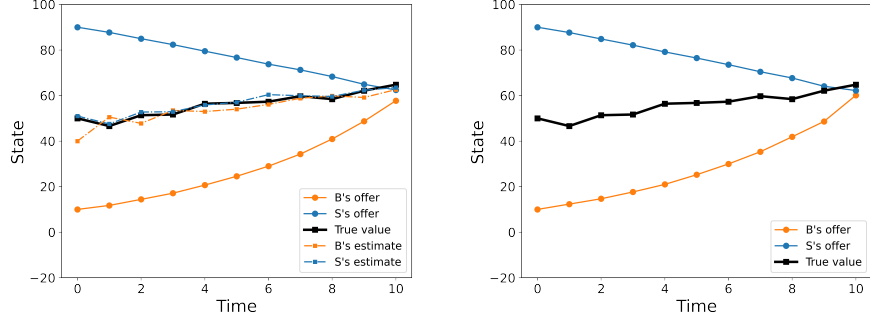


Figure 3: Comparison between the full observation (right) and partial observation (left) cases.

Effect of Information Corrections. With information corrections the buyer's estimate of the value is less affected by the noisy observations (see Figure 4) and players are more likely to achieve an agreement in each of the asymmetric and symmetric case (see Table 5). Compared with the case when $n_1 = 1$, the gap between the number of agreements achieved in the asymmetric and symmetric (inaccurate) case is larger, since players receive noisy signals from the different factors.

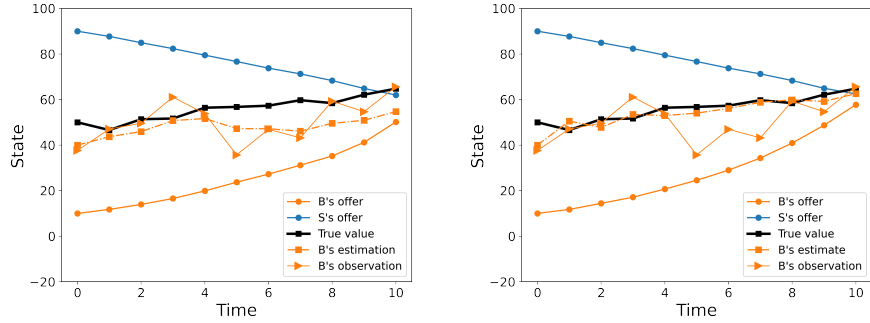


Figure 4: The buyer's price estimate with information corrections (right) and without information corrections (left).

	Asymmetric	Symmetric ("accurate")	Symmetric (inaccurate)
With IC	329	439	315
Without IC	237	440	207

Table 5: Number of agreements achieved in 500 experiments with and without information corrections (IC).