Approximating solution of stochastic differential games for distributed control of a water network

Misra, Rahul; Wisniewski, Rafal; Kallesøe, Carsten S.

Published in:
IFAC-PapersOnLine

DOI (link to publication from Publisher):
10.1016/j.ifacol.2022.09.009

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Misra, R., Wisniewski, R., & Kallesøe, C. S. (2022). Approximating solution of stochastic differential games for distributed control of a water network. *IFAC-PapersOnLine*, *55*(16), 110-115. https://doi.org/10.1016/j.ifacol.2022.09.009

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-16 (2022) 110–115

Approximating solution of stochastic differential games for distributed control of a water network ⋆

Rahul Misra* Rafał Wisniewski* Carsten S. Kallesøe*,**

* *Automation and Control section, Department of electronic systems, Aalborg University, Aalborg, Denmark (e-mail: {rmi, raf, csk}@es.aau.dk).*

** *Grundfos A/S, Poul Due Jensens Vej, Bjerringbro, Denmark (e-mail: ckallesoe@grundfos.com).*

Abstract: In this paper, our objective is to design a distributed optimal control for pumping stations operating in a Water Distribution Network (WDN), where we would like to satisfy consumer demands with minimum energy consumption. The WDN has been modeled using graph theory and stochastic differential equations. This leads to a non-zero sum stochastic differential game. We have approximated the solution of the aforementioned game using Markov chain approximation and combined it with Shapley's algorithm so as to obtain Minimax mixed strategies. Minimax solution can be obtained as a distributed computation at the pumping stations without any knowledge of the costs incurred by the other pumping stations. Simulation results on the water network show convergence to an approximate Minimax solution.

Keywords: Stochastic Games, Stochastic differential equations, Dynamic programming, Markov chain approximation, Water distribution networks.

1. INTRODUCTION

Efficient pressure management in a Water Distribution Network (WDN) is a complex control problem since it entails an inherently multi-input, multi-output system with control objectives of ensuring supply with minimal variance in pressure at demand side while ensuring energy efficiency of supply pumps. The WDN considered in this work consists of suppliers and consumers which are connected together by a piping network. Such a WDN can be modeled using graph theory which represents the topology of the network connecting individual components as stated in Tahavori et al. (2012). The uncertainty in consumption pattern is the reason for the stochastic nature of a water distribution network and therefore, we have used Stochastic Differential Equations (SDE) where the diffusion matrix takes into account the uncertainty due to demand side consumption.

For solving stochastic control problems, Dynamic Programming (DP) is one of the fundamental mathematical tools and forms basis of Reinforcement learning (RL) algorithms (see Bertsekas (2018) for an introduction to dynamic programming). A generalization of DP to include multiple controllers called *Stochastic Games* was introduced in Shapley (1953). The survey paper Raghavan and Filar (1991) and the book Filar and Vrieze (2012) provide an overview of development in the field of Stochastic Games. All the aforementioned settings have a finite number of states which makes it possible to represent the value of each state in a tabular representation. Bellman's optimality equation (or Shapley's equation in the case of Stochastic Games) can than be used to find the optimal states and the corresponding sequence of control actions which constitute the optimal control policy. However, in a practical setting such as a robot, satellite, HVAC system or a water distribution network, the dynamics are modeled using differential equations derived from physical laws such as mass or energy conservation. The state space of such equations is infinite, thereby making the aforementioned algorithms not applicable in such settings. To overcome this limitation, function approximation based techniques such as least squares or a neural networks are used for approximating the value function or the control strategy or both simultaneously. Function approximation based RL algorithms come with some inherent challenges such as non-convergence to the target (see chapters 9, 11 and references therein from Sutton and Barto (2018)) and non-convergence to the saddle point in the case of linear quadratic zero sum dynamic games (see Mazumdar et al. (2020)).

The method of Markov chain approximation (MCA) was developed in Kushner and Dupuis (2001) as a numerical method for solving stochastic control problems and can be considered as an alternative to standard functional approximation techniques. This method has been extended to stochastic differential games and convergence of the value to underlying stochastic process has been proved in Kushner (2002) for zero sum games and in Kushner (2007) for non-zero sum games. MCA has been applied for study of cooperative Nash equilibria in fishery games in Haurie et al. (1994). In this work, we have considered the model based setting and define cost functions which result in a

⋆ This work is supported by Poul Due Jensen Foundation (Grundfos Foundation) under the project SWIfT (Smart water infrastructure).

10.1016/j.ifacol.2022.09.009

non-zero sum differential game. We approximate the solution of the aforementioned game using MCA and Minimax strategies. The reason behind using Minimax strategies for a non zero sum game is to ensure distributed computation of control strategies with private cost matrices, as the pumping stations in fig. 1 are located geographically far away. These methods can also be extended to the model free RL setting as described in Munos and Bourgine (1997). The key contributions of this work are summarized as follows:

- MCA is applied on a SDE model of a WDN in the setting of a non-zero sum differential game.
- A distributed control algorithm based on Shapley's algorithm and MCA is simulated on the WDN.

The rest of the paper is organized as follows. In the next Section, we describe the WDN as an SDE. In Section 3, we introduce MCA method for stochastic control. In Section 4, we introduce stochastic games and present an algorithm combined with MCA introduced in Section 3. In Section 5 we present the simulation results. Lastly, we present our conclusions and future work.

2. WDN REPRESENTED AS AN SDE

A general model of the WDN with $\mathcal{N}$ controllers is defined by state equations of the following type

$$dx = \left(\mathrm{f}(x) + \sum_{k=1}^{\mathcal{N}} \mathrm{g}^k(x)u^k\right)dt + \sigma(x)d\mathrm{w}, \qquad (1)$$

where $x \in X \subseteq \mathbb{R}^n$ is a stochastic process which represents the free flows in the WDN, $u \in U \subseteq \mathbb{R}^p$ represents the pressure control action due to pumps, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^p$ represent the drift, $\sigma : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^l$ represents the diffusion with $a(\cdot) = \sigma(\cdot)\sigma(\cdot)^T$ being the corresponding diffusion matrix, w represents the standard Wiener process on $\mathbb{R}^l$. For simplicity, we shall consider the WDN network shown in fig. 1 with 2 controllers (2 pumping stations) although, the methods developed in this paper are extendable to complex controlled systems with more control inputs. The aim of the pumping stations 1

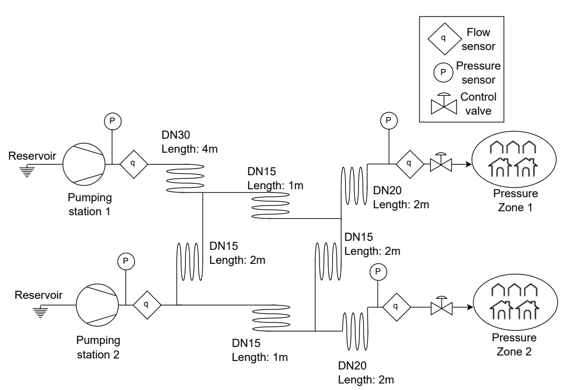

Fig. 1. Process and Instrumentation diagram

and 2 in fig. 1 is to minimize the pressure variation at pressure zones 1 and 2. The WDN is modeled by a directed graph $\Gamma = \{N, E\}$ in fig. 2 where $N = \{n_1, \cdots, n_{11}\}$ represents the nodes or vertices and $E = \{e_1, \cdots, e_{11}\}$ represents the edges where components of WDN such as pipes, pumps and valves are connected. The pumps in fig. 1 are represented as edges e_{10} and e_{11} in fig. 2 and the demands are represented by edges e_8 and e_9. The model

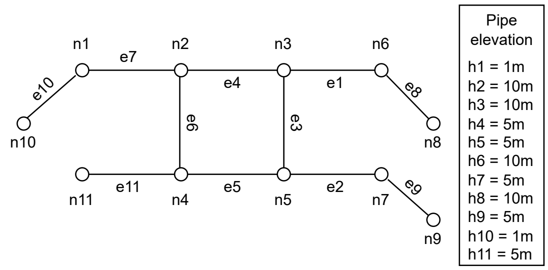

Fig. 2. Graph of the WDN in fig. 1 with pipe elevation $\bar{z}$.

is defined with following assumptions.

Assumption 2.1. The graph Γ is a connected graph.

Assumption 2.2. We require existence and uniqueness of solutions of (1). Let Ω represent the sample space, $\mathcal{F}$ represent the event space and P represent the probability function. Strong existence holds if for a given probability space $(\Omega, \mathcal{F}, P)$, a filtration $\mathcal{F}_t$, an $\mathcal{F}_t$-Wiener process w and an $\mathcal{F}_0$- measurable initial condition $x(0)$, there exists an $\mathcal{F}_t$-adapted process $x(t)$ satisfying (1) for all $t \geq 0$. Furthermore, uniqueness holds if for any two sample paths $x_1(t), x_2(t)$, $P\{x_1(0) = x_2(0)\} = 1 \implies P\{x_1(t) = x_2(t) \forall t \geq 0\} = 1$.

The model of WDN is derived in Misra et al. (2022) and due to space constraints we refer the reader to the same. The stochastic nature of water consumption at nodes n_8 and n_9 in fig. 2 is captured as a Wiener process.

Assumption 2.3. The diagonal terms of diffusion matrix are dominant over off-diagonal terms as follows

$$a_{ii}(x) - \sum_{j:j\neq i} |a_{ij}(x)| \geq 0. \qquad (2)$$

This assumption is valid for WDN as in practice only the free flows at end-user edges (e_8 and e_9 in fig. 2) are correlated with each other. The correlation between free-flows in the other edges are accounted for in the drift term (1). Therefore, the off-diagonal terms are negligible relative to diagonal terms and can be ignored. Due to uncertain water demands and correspondingly stochastic nature of free flows in (1), the pressure measurement is represented via an Ito integral (reviewed in Kushner and Dupuis (2001)) with h representing the stochastic integrand. We assume that h is a right-continuous, adapted and a locally bounded process. We approximate the Ito integral using a mesh m with grid size $\mathcal{Y}_m \to 0$ as,

$$y = \int_s^t \mathrm{h}dx = \lim_{m\to\infty} \sum_{[t_{i-1},t]\in\mathcal{Y}_m} \mathrm{h}_{t_{i-1}}(x_{t_i} - x_{t_{i-1}}). \qquad (3)$$

We shall now define the control objectives of both the players as the following stage cost functions for a player k,

$$c^k(y, x, u) = y^T W_1 y + W_2 \left|xu^k\right|, \qquad (4)$$

where $|\cdot|$ represents the standard 1-norm, W_1 and W_2 are normalized weights. The first term in (4) represents the pressure variance at consumer nodes and the second term represents the energy consumption (kW) for a pumping station k.

3. MARKOV CHAIN APPROXIMATION

In Kushner and Dupuis (2001), the authors have presented a numerical method referred to as Markov chain approximation (MCA) for solving stochastic control problems. MCA finds the reachable states under a given control policy and calculates transition probabilities via application of finite differences method on the Hamilton-Jacobi equation. At any time s, the number of reachable states depends on the time step ahead (i.e. interpolation time) and MCA method automatically calculates the interpolation time (see fig. 3 for an illustration of MCA). If the dynamics

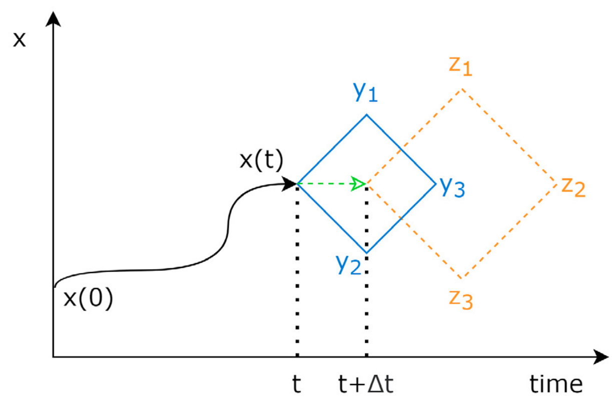

Fig. 3. Here we illustrate MCA by a sample path of (1) starting at $x(0)$ and reaching $x(t)$ at time t, y_1, y_2 and y_3 represent the reachable states for $x(t)$ in time interval Δt. Any realized state $x(t{+}\Delta t)$ (shown by the green arrow) can be found as a convex combination of extremities of blue polygon. Once the next state $x(t+\Delta t)$ is reached the process repeats as shown by orange polygon with reachable states z_1, z_2 and z_3.

and/or diffusion are large in magnitude at a given time, the interpolation time correspondingly becomes smaller. The calculated probabilities are a function of the dynamics alone and therefore this method can be used to represent (1) as transition probabilities. We consider $\mathcal{B} \subset \mathbb{R}^n$ as a compact state space with absorption on boundary. The discounted stochastic control problem for a player k with $\tau = \inf\{t : x(t) \notin \mathcal{B}\}$ is the minimum time taken to exit $\mathcal{B}$ can now be defined as follows.

$$\inf_{u^k} \mathop{\mathbb{E}}_{x(s),u(\cdot)} \left(\int_s^\tau \gamma^t c^k\left(x(t), u(t)\right) dt + \gamma^\tau c^k_\tau\left(x(\tau)\right) \right) \quad (5a)$$

$$\text{s.t.} dx = \left(\text{f}(x(t)) + \sum_{k=1}^{\mathcal{N}} \text{g}^k(x(t)) u^k(t) \right) dt + \sigma d\text{w}, \quad (5b)$$

$$u^k(t) \in U, \quad x(t) \in \mathcal{B}, \quad (5c)$$

where $x(s)$ is the state at some starting time s, u is a vector with all $\mathcal{N}$ players control actions, c^k is running cost for player k, c^k_τ is the terminal cost for player k and γ is the discount factor. We have assumed that the control space U is the same for all the players for simplicity although the methods discussed in this work are easily applicable to problems where control spaces are different. Let the stochastic integral in the objective function of (5) be denoted by

$$Q^k_s(x, u^k) = \mathop{\mathbb{E}}_{x(s),u(\cdot)} \left(\int_s^\tau \gamma^t c^k\left(x(t), u(t)\right) dt + \gamma^\tau c^k_\tau\left(x(\tau)\right) \right), \quad (6)$$

where subscript s indicates dependence on s. The optimal value can than be obtained by minimizing Q^k_s with respect to u^k for each state as follows

$$V^k_s(x) = \inf_{u^k} Q^k_s(x(s), u^k). \quad (7)$$

Note that the above formulation gives rise to a non-zero sum differential game. Since the value of a state for any player (7) depends on the joint action u of the all the players, we need to define the solution in the sense of strategies π of all the players. In the sequel, the notation $-k$ shall denote all the players except player k and let each player choose control action u^k from probability distribution π^k. The approximate Minimax solution for (5) can be defined as the mixed strategy π^{k*} such that,

$$Q^k_s(x, \pi^{k*}, \pi^{-k}) - \epsilon \leq Q^k_s(x, \pi^{k*}, \pi^{-k*}) \leq Q^k_s(x, \pi^k, \pi^{-k*}) + \epsilon, \quad (8)$$

where ϵ indicates the amount of tolerable sub-optimality to the exact Minimax solution. For discretizing (5), we divide the compact state space $\mathcal{B}$ into v number of discrete states. Consequently, we can define the state space approximation parameter h (which represents the coarseness of the state space grid) as follows

$$h = \frac{1}{v-1}. \quad (9)$$

Let $\xi(s) \in \mathbb{R}^v$ denote the discrete state centered at $x(s)$ with a grid box of dimension h at some time s, than the next reachable discrete state in the set $\mathcal{B}$ is denoted by $\xi(s+1) = x(s) + e_i h$, where $e_i \in \mathbb{R}^n$ is the basis vector in i^{th} direction of the state space (see fig. 4). Note that we are

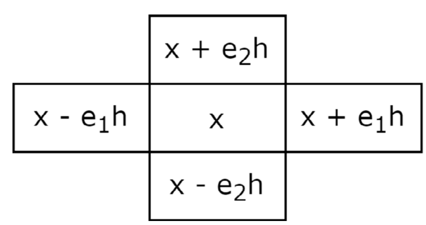

Fig. 4. Illustration of state space as a grid. Any state whose numerical value is within the boundary of a particular grid box is considered as a part of the same.

considering only state transitions of the type $x \pm e_i h$ and not of the type $x \pm e_i h \pm e_j h$ since the interpolation time is small (depending on the drift and diffusion as we shall see later) and due to Assumption 2.3. These additional terms can be considered at the cost of a greater computation time. We shall now apply the well-known Ito's lemma in order to obtain evolution of $V^k_s(x)$ with time. In Kushner (2007), convergence of Markov chain approximation for non-zero sum differential games is proved based on certain key assumptions.

Assumption 3.1. (Kushner (2007)). The drift of (1) is additively separable into two components which represents the contributions to the drift due to individual players. Furthermore, the cost c^k for each player is also additively separable into two components which represents the contributions to the cost due to the individual players.

Assumption 3.1 is satisfied for both the considered system (1) and the cost function (4) and can be verified by dividing the drift in (1) as

$$b^k(x,u) = \frac{\mathrm{f}(x)}{\mathcal{N}} + \mathrm{g}^k(x)u^k. \tag{10}$$

Similarly (4) can be split among the players. Let $a_{ij}(x)$, $i,j = 1,\cdots,n$ be an element of the diffusion matrix $a(x)$ and let $b_i(x,u)$ represent the drift term $\mathrm{f}_i(x) + \sum_{k=1}^{\mathcal{N}} \mathrm{g}^k(x)u^k$, than we can define $\mathcal{L}$ for (1) as follows.

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x(t))\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x(t),u(t))\frac{\partial}{\partial x_i}. \tag{11}$$

Applying Ito's differential operator on (5) gives us the stochastic analogue of the well-known Hamilton-Jacobi equations (see Kushner and Dupuis (2001)) for a player k as follows.

$$\begin{aligned} &\mathcal{L}V^k(x(t)) + \gamma^t c^k(x(t),u(t)) = 0,\\ &V_s^k(x(\tau)) = c^k(x(\tau)), V_s^k(x(s)) = c^k(x(s)), \end{aligned} \tag{12}$$

where the last two equations represents the boundary conditions. Assume that each player acts simultaneously at a given time instant to ensure well-posedness of solutions of (12). Expanding the differential operator in (12) gives,

$$\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 V_s^k(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x,u)\frac{\partial V_s^k(x)}{\partial x_i} + \gamma^t c^k(x,u) = 0, \tag{13}$$

where we have written $x(t), u(t)$ as x, u for notational convenience. We now apply finite-differences with space approximation parameter h on (13) as follows.

$$\begin{aligned} &\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{V_s^k(x+e_ih) + V_s^k(x-e_ih) - 2V_s^k(x)}{h^2}\\ &+ \sum_{i=1}^{n} |b_i(x,u)|^+ \frac{V_s^k(x+e_ih) - V_s^k(x)}{h}\\ &- \sum_{i=1}^{n} |b_i(x,u)|^- \frac{V_s^k(x) - V_s^k(x-e_ih)}{h} + \gamma^t c^k(x,u) = 0, \end{aligned} \tag{14}$$

where $|b_i(x,u)|^+ = \max(b_i(x,u),0)$ and $|b_i(x,u)|^- = \max(-b_i(x,u),0)$. This is the standard "upwind" scheme in numerical analysis of hyperbolic partial differential equations (see LeVeque (2007)). The intuition behind this scheme is that the approximation of V_s^k should be in the same direction as the drift of (1). It can also be verified that $|b_i(x,u)|^+ + |b_i(x,u)|^- = |b_i(x,u)|$. In order to make the probability transitions independent of joint control action u, we calculate the joint control action u_{max} which gives maximal drift and replace $|b_i(x,u)|$ with $|b_i(x,u_{max})|$. We can rearrange the terms in (14) so as to obtain elements of probability transition matrix $p(x, x \pm e_ih \mid u)$ and interpolation time Δt as follows.

$$\begin{aligned} V_s^k(x) = &\underbrace{\frac{a_{ii}(x)/2 + h\,|b_i(x,u)|^+}{\sum_{i=1}^{n}(a_{ii}(x) + h\,|b_i(x,u_{max})|)}}_{p(x,x+e_ih|u)} V_s^k(x+e_ih)\\ &+ \underbrace{\frac{a_{ii}(x)/2 + h\,|b_i(x,u)|^-}{\sum_{i=1}^{n}(a_{ii}(x) + h\,|b_i(x,u_{max})|)}}_{p(x,x-e_ih|u)} V_s^k(x-e_ih)\\ &+ \underbrace{\frac{h^2}{\sum_{i=1}^{n}(a_{ii}(x) + h\,|b_i(x,u_{max})|)}}_{\Delta t} \gamma^t c^k(x,u). \end{aligned} \tag{15}$$

The probability of state remaining unchanged is given by

$$p(x,x \mid u) = 1 - (p(x,x+e_ih \mid u) + p(x,x-e_ih \mid u)). \tag{16}$$

It should be noted that we only get the transition probabilities required for the next state transition and therefore, the probability transition matrix constructed using this method will be sub-stochastic (i.e. sum of probabilities will be less than 1). Thus, the differential game (5) can be approximated by a Stochastic Game in the sense of Shapley (1953).

4. SOLVING STOCHASTIC GAMES

We are proposing a distributed Stochastic Game solver based on Shapley's algorithm which can be solved by each of the players without any knowledge of the cost incurred by other player. Stochastic games are a generalization of static games where the decisions taken by a player influences both the immediate costs and also the states reached in the future. Formally a stochastic game $\mathcal{G}$ can be defined as a tuple $\mathcal{G} = \{\mathcal{N}, \mathcal{X}, U, \mathcal{P}, (C^1,\cdots,C^{\mathcal{N}}), \gamma\}$, where $\mathcal{N}$ is the no. of players, $\mathcal{X}$ is the finite state space of the game, U is the finite control space of all the players, $\mathcal{P}$ is the probability transition matrix, $C^1,\cdots,C^{\mathcal{N}}$ are the cost matrices for each player corresponding to U and γ is the discount factor. We begin by discretizing the control space U into finite control actions u. For the considered WDN, these represent the operating power of pumping stations (for ex. $u^1 = 1$ implies that the pumping station 1 is operating at 10% of its maximum capacity).

Assumption 4.1. A player only knows the possible finite control actions that can be taken by the other players.

Let $C^k \in \mathbb{R}^v \times \mathbb{R}^{u^1} \times \cdots \times \mathbb{R}^{u^{\mathcal{N}}}$ denote the linear operator representing cost for a player k at a discrete state $\xi(s)$. For the considered 2-player game, C^k has the dimensions $\mathbb{R}^{v \times u^1 \times u^2}$. Every non-zero sum game $\mathcal{G} = \{\mathcal{N}, \mathcal{X}, U, \mathcal{P}, (C^1,\cdots,C^{\mathcal{N}}), \gamma\}$ can be solved using Minimax strategies if each player k solves the corresponding zero-sum game $\mathcal{G}' = \{\mathcal{N}, \mathcal{X}, U, \mathcal{P}, (C^k, -C^k), \gamma\}$ to obtain their worst-case costs. Such solutions are referred to as mixed security strategies in Alpcan and Başar (2010). In the sequel $\xi = \xi(s) \in \mathcal{X}$ shall denote a finite state of the game $\mathcal{G}$ at time s and the operator $val[\cdot]$ denotes the value V of a matrix game as per Von Neumann's Minimax theorem which states that for any real $n_1 \times n_2$ matrix A with elements a_{ij}, there exists a pair of probability vectors $\pi^{\mathbf{1}*} = (\pi_1^{1*},\cdots,\pi_{n_1}^{1*})$ and $\pi^{\mathbf{2}*} = (\pi_1^{2*},\cdots,\pi_{n_2}^{2*})$ such that

$$\sum_i a_{ij}\pi_i^{1*} \leq V \leq \sum_j a_{ij}\pi_i^{2*} \quad \forall i, j. \tag{17}$$

Let $u = (u^1, \cdots, u^{\mathcal{N}})$ be the joint control actions and v be the number of discrete states. A Stochastic game encodes the transition probabilities in the cost matrix by defining Shapley game matrix as follows.

$$M^k(\xi, u) = c^k(\xi, u) + \gamma \sum_{\chi=1}^{v} p(\chi|\xi, u) V_s^k(\chi), \tag{18}$$

where $c^k(\xi, u) \in C^k$ is the cost at state ξ, γ is the discount factor, χ represents the reachable states from ξ, $p(\chi|\xi, u)$ is the transition probability of reaching χ given ξ and $V_s^k(\chi)$ is the value of state χ. The mixed security strategy $\pi^k(\xi)$ for the game $\mathcal{G}$ can be computed efficiently using the following linear program.

$$\min_{\pi^k} \; V_s^k(\xi) \tag{19a}$$

$$\text{s.t.} \quad \sum_{u^k \in U} M^k(\xi, u)\pi^k \leq V_s^k(\xi), \forall u^{-k} \in U, \tag{19b}$$

$$\sum_{u^k} \pi^k = 1, \tag{19c}$$

$$\pi^k \geq 0, \; \forall u^k \in U, \tag{19d}$$

where constraint (19b) ensures that player k chooses mixed strategy π^k such that for every joint pure strategy profile $u^{-k} \in U$ of other players, player k's expected value for state ξ is at most $V_s^k(\xi)$ (which is being minimized in (19a)) and remaining constraints ensure that mixed strategy π^k obeys axioms of probability. We now state the MCA based algorithm for solving stochastic games in Algorithm 1. The value of the game $\mathcal{G}$ is the unique solution ϕ of the system $\phi^k = val[M^k(\phi)]$ for all $k = 1, \cdots, \mathcal{N}$ as proved in Shapley (1953).

Algorithm 1 MCA Stochastic differential game solver

Input: Initial state $x(s)$, State space Grid $\Delta^h x$, drift b, diffusion matrix a, control space U, terminal time T.
Initialize $V_s^k = 0$ for all states in the grid
while $s < T$ **do**
  $\xi \leftarrow x(s)$
  Find grid square $\Delta^h x(s)$ corresponding to ξ
  Construct cost matrix C^k, $\forall u^1, \cdots, u^{\mathcal{N}}$
  Let χ denote possible reachable states from ξ
  Obtain transition probabilities using (15) and (16)
  for All possible control actions of players $-k$ **do**

$$V_s^k(\xi) \leftarrow val\Bigg[\underbrace{c^k + \gamma \sum_{\chi=1}^{v} p(\chi|\xi, u) V_s^k(\chi)}_{M^k}\Bigg] \tag{20}$$

  end for
  Solve the Shapley game (20) using (19) for $\pi^k(\xi)$
  Sample u^k from $\pi^k(\xi)$ and apply on (1) at time s
  $s \leftarrow s + \Delta t$
end while

5. SIMULATION RESULTS AND DISCUSSIONS

We ran a variety of simulations with different initial conditions on (1). The simulation results presented in this section are with initial conditions $x = [-50\ 60\ 50\ 0]^T$ and with $v = 20$ discrete states. In fig. 5, we can observe that

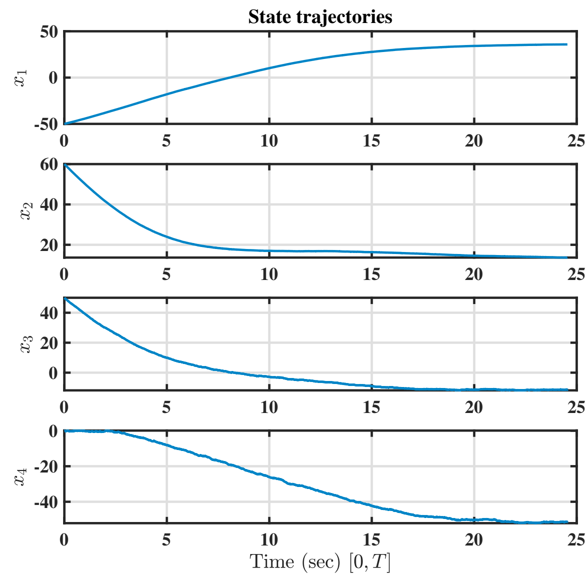

Fig. 5. State trajectory for initial $x = [-50\ 60\ 50\ 0]^T$

the states representing the free flows in the network reach a steady state and the system dynamics are stable. In fig.

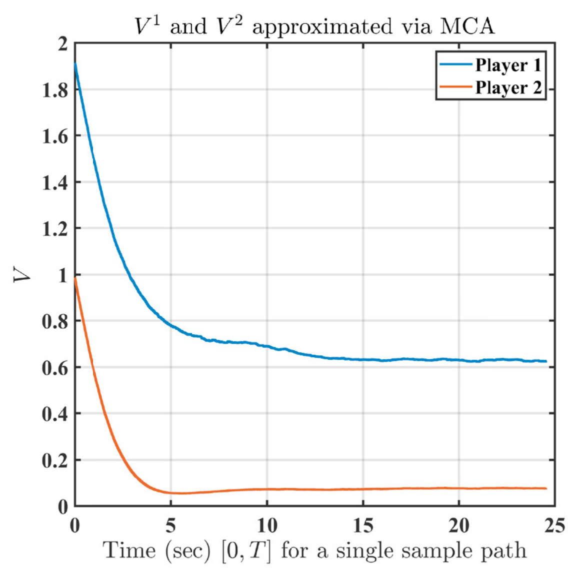

Fig. 6. The values obtained for both the players converge to the Minimax solution. The state space discretization was $v = 20$.

6, we can observe that the optimal values for both the players reach a fixed point of the Shapley equation (20). The value functions of both the players decrease to their minimum as both the players apply control actions. Player 1 applies higher control input relative to Player 2 and consequently Player 1's value is higher than Player 2. This situation can be changed by modifying the weights W_1 and W_2 in (4). The accuracy of Algorithm 1 is dependent on the coarseness of state grid which is determined by the considered number of discrete states v. In table 1, we show how the numerical accuracy improves with higher

v. It is proved in Kushner (2002) and Kushner (2007)

No. of discrete states	V^1 obtained from MCA	V^2 obtained from MCA
$v = 5$	13.2525	1.7103
$v = 10$	7.8972	0.9951
$v = 15$	1.5339	0.1712
$v = 20$	0.6549	0.0732

Table 1. Value function approximations obtained at terminal time T for different v.

that the Minimax solution of original continuous time problem (5) can be approximated upto an arbitrary $\epsilon > 0$ by making the grid parameter h small (which increases simulation time in practice) and this is shown in the table 1. In fig. 7, we can observe that the output y successfully tracks the reference pressure. The spikes in fig. 7 are

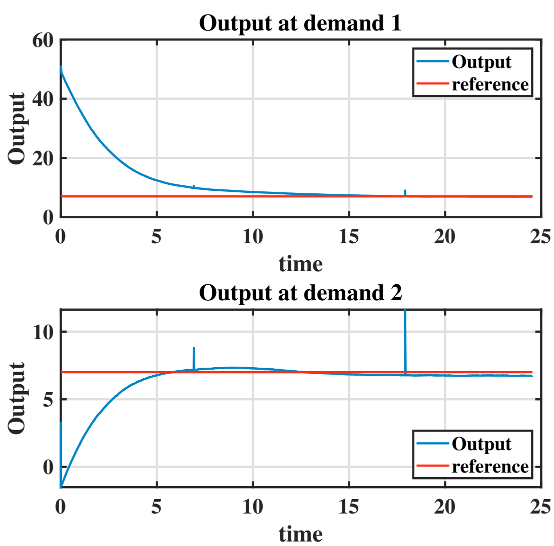

Fig. 7. Output pressure is tracked successfully to the reference.

due to the controllers periodically switching off as there is a cost associated with operation in (4). A Minimax solution may or may not always converge to an Nash equilibrium solution unless the non-zero sum differential game is strategically equivalent to a zero-sum differential game (see Başar and Olsder (1998)). However, in this case, simulation studies show that the Minimax solution does correspond to the Nash equilibrium solution.

6. CONCLUSION

We have designed a control strategy for (5) using MCA and Stochastic games. We have obtained approximate Minimax strategies and simulation results show that steady state is reached. However, better solutions can be obtained if the players have access to the cost matrices or control actions of players at previous iteration. Given the challenging nature of non zero sum differential games and stochastic games in general (see Bressan (2011) and Filar and Vrieze (2012)), we do obtain a reasonable solution by using Minimax strategies and MCA. The Minimax solution can also be extended to $\mathcal{N} > 2$ players in a straight forward as each player only needs to know their own cost matrix. However, the algorithm scales exponentially as the number of players is increased. Besides scalability in future, we would also like to consider model-free learning of Nash equilibrium in this setting.

ACKNOWLEDGEMENTS

Financial support from the Poul Due Jensen Foundation (Grundfos Foundation) for this research is gratefully acknowledged.

REFERENCES

Alpcan, T. and Başar, T. (2010). *Network security: A decision and game-theoretic approach*. Cambridge University Press.

Başar, T. and Olsder, G.J. (1998). *Dynamic noncooperative game theory*. SIAM.

Bertsekas, D.P. (2018). *Abstract dynamic programming*. Athena Scientific.

Bressan, A. (2011). Noncooperative differential games. *Milan Journal of Mathematics*, 79(2), 357–427.

Filar, J. and Vrieze, K. (2012). *Competitive Markov decision processes*. Springer Science & Business Media.

Haurie, A., Krawczyk, J.B., and Roche, M. (1994). Monitoring cooperative equilibria in a stochastic differential game. *Journal of optimization Theory and Applications*, 81(1), 73–95.

Kushner, H.J. (2002). Numerical approximations for stochastic differential games. *SIAM journal on control and optimization*, 41(2), 457–486.

Kushner, H.J. (2007). Numerical approximations for nonzero-sum stochastic differential games. *SIAM journal on control and optimization*, 46(6), 1942–1971.

Kushner, H.J. and Dupuis, P.G. (2001). *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Science & Business Media.

LeVeque, R.J. (2007). *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*. SIAM.

Mazumdar, E., Ratliff, L.J., and Sastry, S.S. (2020). On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1), 103–131.

Misra, R., Wisniewski, R., and Kallesøe, C.S. (2022). Approximating the model of a water distribution network as a markov decision process. In *10th Vienna Conference on Mathematical Modelling*.

Munos, R. and Bourgine, P. (1997). Reinforcement learning for continuous stochastic control problems. In *NIPS*, 1029–1035.

Raghavan, T. and Filar, J.A. (1991). Algorithms for stochastic games—a survey. *Zeitschrift für Operations Research*, 35(6), 437–472.

Shapley, L.S. (1953). Stochastic games. *Proceedings of the national academy of sciences*, 39(10), 1095–1100.

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tahavori, M., Kallesøe, C.S., Leth, J., and Wisniewski, R. (2012). Modeling of water supply systems: Circuit theoretic approach. In *2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 1561–1566. IEEE.