Working Waper manuscript No.
(will be inserted by the editor)

Approximating Solutions for Nonlinear Dynamic Tracking Games

Doris A. Behrens · Reinhard Neck

Accepted: 23 January 2014

Abstract This paper presents the OPTGAME algorithm developed to iteratively approximate equilibrium solutions of 'tracking games', i.e. discrete-time nonzero-sum dynamic games with a finite number of players who face quadratic objective functions. Such a tracking game describes the behavior of decision makers who act upon a nonlinear discrete-time dynamical system, and who aim at minimizing the deviations from individually desirable paths of multiple states over a joint finite planning horizon. Among the noncooperative solution concepts, the OPTGAME algorithm approximates feedback Nash and Stackelberg equilibrium solutions, and the open-loop Nash solution, and the cooperative Pareto-optimal solution.

We thank five anonymous referees; the editor; and Engelbert Dockner for fruitful discussions; Christina Kopetzky for indefatigable organizational support; and delegates of numerous workshops where earlier versions of this paper were presented for inspiring comments. We gratefully acknowledge financial support from the EU Commission (project no. MRTN-CT-2006-034270 COMISEF), the Austrian Science Foundation (FWF-project no. P12745-OEK), the Austrian National Bank (OeNB-projects no. 9152, no. 11216, and no. 12166), the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF-grant no. 20214/23793/35529), the Carinthian Chamber of Commerce and the Provincial Government of Carinthia, and the former Ludwig Boltzmann Institute for Economic Policy Analyses, Vienna.

Doris A. Behrens
Alpen-Adria Universität Klagenfurt, Department of Controlling and Strategic Management, Universitätsstr. 65-67, 9020 Klagenfurt/Wörthersee, Austria
Lakeside Labs GmbH, 9020 Klagenfurt/Wörthersee, Austria
Vienna University of Technology, Department of Mathematical Methods in Economics, Argentinierstr. 8, 1040 Vienna, Austria
E-mail: doris.behrens@aau.at

Reinhard Neck
Alpen-Adria Universität Klagenfurt, Department of Economics, Universitätsstr. 65-67, 9020 Klagenfurt/Wörthersee, Austria
E-mail: reinhard.neck@aau.at

Keywords Dynamic game theory · Algorithm · Noncooperative game · Cooperative game · Tracking problem · Nash equilibrium · Markov-perfect equilibrium · Stackelberg equilibrium · Pareto optimum

1 Introduction

Since the turn of the century, we have seen various applications of the theory of dynamic games [1], which is appropriate for modeling policy coordination problems in, e.g., business and management sciences (e.g. [2]), industrial organization (e.g. [3,4]), advertising (e.g. [5]), marketing (e.g. [6]), economic policy design (e.g. [7,8,9,10,11,12]), the economics of natural resources (e.g. [13]), and even for explaining undesirable social phenomena like the upsurge of terrorism (e.g. [14,15,16]). Nonetheless, due to the complex nature of a multi-decision-maker conflict situation, with the exception of some useful yet narrow classes,[1] computational algorithms that provide numerical solutions to game-theoretic problems are rare. Pioneering (and partially ongoing) work does exist, but faces significant limitations with respect to the structure of the problem that it can handle.[2] These restrictions pertain, in particular, to the way nonlinearity is addressed, the degree to which the players' payoffs are perceived to be related, and the choice of the time structure.

In this paper, we present a computational algorithm which is designed specifically for economic policy applications and allows for the approximation of equilibrium solutions of discrete-time deterministic dynamic games whose constraints are given by a system of nonlinear difference equations in a particular state space description. In particular, we consider the evolution of choices made over time by a finite number of decision makers who aim at minimizing deviations from individually desirable trajectories of the system, thus playing a 'tracking game'. Such a game is an extension of the linear regulator problem, a single decision maker's tracking problem that is well known from LQ optimal control theory (e.g. [17,26,27]). For economic applications, especially the Stackelberg game provides a microeconomic foundation for *ad hoc* macroeconomic models with forward-looking behavior [28,7]. The latter is important,

[1] Linear quadratic dynamic games (LQDG) are, for example, a model class that is very well understood and extensively investigated, especially in continuous time [17]. For a recent piece of work applying deterministic, finite-dimensional, zero-sum LQDGs formulated in discrete time, see, e.g. Pachter and Pham [18].

[2] The computer program DYNGAME [19] is, for example, designed to solve rational expectations models and related deterministic dynamic game problems for economic applications, but is not appropriate for approximating solutions for nonlinear problems because the algorithm linearizes the nonlinear dynamic constraint prior to initializing the solution procedure. The LQDG Toolbox [20] exclusively solves linear quadratic (LQ) open-loop deterministic dynamic games (DG) in continuous time and does not provide feedback solutions. Regarding dynamic stochastic games, the Pakes & McGuire computer algorithm [21,22,23], which calculates Markov-perfect equilibrium solutions, is set up to investigate dynamic industries with heterogenous firms, i.e. the dynamic competition in an oligopolistic industry with investment, entry, and exit [24]. Here, intensive work is ongoing with respect to both theoretical and computational aspects (see, e.g. [3,4,25]), with potential future applicability to (macro)economic problems such as those for which OPTGAME was developed.

because it explicitly takes into account the reaction of decision makers (governments, private sector agents, etc.) to deliberate changes in economic policy measures by other decision makers, and, thereby, can ensure that policies derived from a model of this class do not suffer from the Lucas [29] critique ([7]).

Going beyond our previous work[3] in terms of the dynamic representation of the constraint, the solution procedure, and, particularly, the number of decision makers, the present version of the OPTGAME algorithm (OPTGAME 2.0) constitutes a computational tool to approximate feedback and open-loop Nash equilibrium solutions, feedback Stackelberg equilibrium solutions, and cooperative Pareto-optimal solutions for deterministic nonzero-sum discrete-time nonlinear quadratic difference games, namely nonlinear tracking games with a finite planning horizon and a finite number of players.

The paper is structured as follows: The class of game-theoretic problems whose solutions can be approximated by the OPTGAME algorithm is defined in Sect. 2.1, while Sect. 2.2 outlines the solution concepts and information patterns. A presentation of how the solutions are approximated follows in Sect. 3, which also contains a detailed description of the iterative procedure for approximating a feedback Nash equilibrium solution (Sect. 3.1). The modifications required for the computation of open-loop Nash, feedback Stackelberg equilibrium, and Pareto-optimal solutions are discussed in Sect. 3.2. Section 4 demonstrates the application of the algorithm in an economic policy example. Section 5 concludes.

2 Nonlinear Quadratic Tracking Games

2.1 The Class of Games Amenable to the OPTGAME Algorithm

The type of intertemporal nonzero-sum game to which our algorithm can be successfully applied models a temporary interaction of a finite number of decision makers who face multiple objectives. With $T \in \mathbb{N}$ indicating the terminal period of a finite planning horizon, we consider the following problem: Each decision maker i $(i = 1, ..., n)$ has to find a control path $\{\mathbf{u}_t^{i*}\}_{t=1}^T$ that minimizes a quadratic payoff functional of the trajectories of the states and controls, i.e.,

$$J^i\big(\{\mathbf{X}_t\}_{t=1}^T\big) = \tfrac{1}{2} \cdot \sum_{t=1}^{T} [\mathbf{X}_t - \tilde{\mathbf{X}}_t^i]' \, \boldsymbol{\Omega}_t^i \, [\mathbf{X}_t - \tilde{\mathbf{X}}_t^i], \tag{1}$$

subject to a constraint to be discussed below (see Eq. 4). For each time period t $(t=1, ..., T)$, $\mathbf{X}_t := [\mathbf{x}_t \;\; \mathbf{u}_t^1 \;\; ... \;\; \mathbf{u}_t^n]' \in \mathbb{R}^m$, where $\mathbf{x}_t \in \mathbb{R}^{m_x}$ is the state vector and $\mathbf{u}_t^i \in \mathcal{U}^i \subseteq \mathbb{R}^{m_i}$ $(i = 1, ..., n)$ with $m := m_x + \sum_{i=1}^n m_i$ is player i's control vector.

[3]It is the number of decision makers involved that distinguishes OPTGAME 2.0 from its predecessor OPTGAME 1.0, which is designed for only two players; see [30,31]. Moreover, the MATLAB implementation of OPTGAME 2.0 offers a wider spectrum of numerical equation solvers than its predecessor.

$\tilde{\mathbf{X}}_t^i := [\,\tilde{\mathbf{x}}_t^i \quad \tilde{\mathbf{u}}_t^{i1} \quad ... \quad \tilde{\mathbf{u}}_t^{in}]'$ with $\tilde{\mathbf{x}}_t^i \in \mathbb{R}^{m_x}$ and $\tilde{\mathbf{u}}_t^{ij} \in \mathbb{R}^{m_i}$ $(i,j = 1,...,n)$ denotes the values of the states and the controls (of any player j) that player i considers as desirable for time period t.

The matrices $\mathbf{\Omega}_t^i \in \mathbb{R}^{m \times m}$ with $\mathbf{\Omega}_t^i = \mathbf{\Omega}_t^{i'}$, are player i's weights of the differences between the actual and the desired values of the target and control variables embodied in the vector $[\mathbf{X}_t - \tilde{\mathbf{X}}_t^i] \in \mathbb{R}^m$. Thus, for all $t \in \{1,...,T\}$, player i's penalty matrices,

$$\mathbf{\Omega}_t^i := \begin{bmatrix} \mathbf{Q}_t^i & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t^{i1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_t^{in} \end{bmatrix}, \tag{2}$$

are of a block-diagonal form, where the symmetric blocks $\mathbf{Q}_t^i \in \mathbb{R}^{m_x \times m_x}$ and $\mathbf{R}_t^{i1} \in \mathbb{R}^{m_1 \times m_1}$, ..., $\mathbf{R}_t^{in} \in \mathbb{R}^{m_n \times m_n}$ contain player i's penalties for deviating from the desired states and from i's desired levels of the controls of player $1,...,n$.[4,5] We require for all $i = 1,...,n$ the matrices $\mathbf{Q}_t^i$, $\mathbf{R}_t^{i1}$, ..., $\mathbf{R}_t^{i(i-1)}$, $\mathbf{R}_t^{i(i+1)}$, ..., $\mathbf{R}_t^{in}$ to be positive semi-definite and the matrices $\mathbf{R}_t^{ii} \in \mathbb{R}^{m_i \times m_i}$ to be positive definite.

In order to determine cooperative solutions, we define a collective payoff function as a convex combination of all players' individual objective functions, $J^i\big(\{\mathbf{X}_t\}_{t=1}^T\big)$, i.e.,

$$J\big(\{\mathbf{X}_t\}_{t=1}^T\big) = \sum_{i=1}^n \mu^i J^i\big(\{\mathbf{X}_t\}_{t=1}^T\big), \qquad \sum_{i=1}^n \mu^i = 1, \tag{3}$$

where weight $\mu^i \in [0,1]$ $(i = 1,...,n)$ reflects player i's 'power', i.e. i's weight in the joint objective value.

The constraints of the decision makers' choices are given by a first-order[6] system of nonlinear autonomous difference equations,

$$\mathbf{x}_t = \mathbf{f}\big(\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{u}_t^1, ..., \mathbf{u}_t^n, \mathbf{z}_t\big), \qquad \mathbf{x}_0 = \bar{\mathbf{x}}_0, \tag{4}$$

with initial state, $\bar{\mathbf{x}}_0 \in \mathbb{R}^{m_x}$. The vector-valued function $\mathbf{f}$ consists of m_x components, $\mathbf{f} = \big[f^1 \; f^2 \; ... \; f^{m_x}\big]'$, and $\mathbf{z}_t \in \mathbb{R}^{m_x}$ is a vector of exogenous variables

[4] Decision makers may also care about what other decision makers do, i.e., it is possible that $\mathbf{R}_t^{ij} \neq \mathbf{0}$ for $j \neq i$. In other words, deviations in the levels of someone else's control variables from what is desired (seen from one's own perspective) can be punished in one's own objective function.

[5] Particularly in economic models, often one accounts for time preferences and assumes the penalty matrices to be $\mathbf{\Omega}_t^i := (\vartheta^i)^{t-1}\mathbf{\Omega}_0^i \;\forall\; t \in \{1,...,T\}$, where $\vartheta^i \in (0,1)$ is each player i's individual discount rate and $\mathbf{\Omega}_0^i \in \mathbb{R}^{m \times m}$ is i's nonnegative definite penalty matrix for the intial time period.

[6] The assumption of a first-order system of difference equations is not restrictive as higher-order difference equations can always be reduced to systems of first-order difference equations by appropriately redefining variables. Lagged state variables can also be easily removed by introducing additional state variables. In OPTGAME, this is done using the procedure proposed in [28, 29].

that are not subject to control by any player. We require the first and second derivatives of $\mathbf{f}$ with respect to $\mathbf{x}_{t-1}, \mathbf{x}_t$, and $\mathbf{u}_t^i \; \forall \; i \in \{1, ..., n\}$ to exist and to be continuous.

The particular representation of the dynamic system given by Eq. 4 frequently appears in economic models [32,33]. It suggests that at the beginning of time period t, when the decision makers decide on their control activities (the policy instruments $\mathbf{u}_t^1, ..., \mathbf{u}_t^n$), only the state vector $\mathbf{x}_{t-1}$ is known, while $\mathbf{x}_t$ is still unknown. Not before the decisions about the policy instruments are made and put into action will they affect the state of the system.

2.2 Solution Concepts and Information Patterns in the OPTGAME Algorithm

Among the solution concepts of dynamic game theory, we consider noncooperative and fully cooperative solution concepts, and ignore intermediate possibilities with coalitions of subsets of players and the question of how cooperation may evolve. For a noncooperative mode of play, we distinguish between Nash equilibriums, where all players act simultaneously, and a Stackelberg equilibrium, where the players act sequentially and assume asymmetric roles as 'leaders' and 'followers'.

A strategy, denoted by φ^i, maps the information set, i.e., the extent of information that is available to player i when making a decision, to the set of i's feasible controls. Based on the information structure, we distinguish between open-loop and feedback (Markov) strategies.

Under an open-loop information pattern, each decision maker can be imagined to commit himself/herself at the beginning of the game to all future actions he/she will take by selecting a time path strategy [34]; see also [17]. Modeling strategies in this way makes sense if none of the players can observe the state variable(s) or deliberately chooses to ignore this information. On the other hand, under a feedback information pattern, the players can be imagined to observe the values of the state variables and react to them by choosing their actions according to a 'decision rule', i.e., $\mathbf{u}_t^i = \varphi_t^i(\mathbf{x}_{t-1}) \; \forall \; t \in \{1, ..., T\}$.

In an (open-loop or feedback) Nash game, the players have symmetric roles. In the feedback Stackelberg game, there is an asymmetry between one leader and one or more followers. At the beginning of time period t, the Stackelberg leader (say, player 1) announces his/her decision rule, $\mathbf{u}_t^1 = \varphi^1(\mathbf{x}_{t-1})$, to the Stackelberg followers (all other players $i = 2, ..., n$), who, in turn, simultaneously respond to the leader's announcement without any coordination among them according to $n-1$ reaction functions, $\mathbf{u}_t^i = \varphi^i(\mathbf{x}_{t-1}, \mathbf{u}_t^1)$. The leader then incorporates the followers' best rational responses to the leader's anouncement $\partial \mathbf{u}_t^i / \partial \mathbf{u}_t^1 \; (i = 2, ..., n)$ into his/her decision.

In addition to the open-loop and feedback Nash equilibrium and feedback Stackelberg equilibrium solutions, OPTGAME also determines the cooperative Pareto-optimal solution.

3 How the OPTGAME Algorithm Works

In addition to the parameters introduced in Sect. 2.1, i.e., n, T (and possibly μ^1, ..., μ^n, $\vartheta^1, ..., \vartheta^n$), and the dynamic system ($\mathbf{f}$, $\mathbf{x}_0$), the user of the OPTGAME algorithm has to provide the entire time path of the exogenous variables $\{\mathbf{z}_t\}_{t=1}^T$. The necessary inputs include, moreover, the time paths for the targets and the penalties for each player i, i.e., $\{\tilde{\mathbf{x}}_t^i\}_{t=1}^T$, $\{\mathbf{Q}_t^i\}_{t=1}^T$, $\{\tilde{\mathbf{u}}_t^{i1}\}_{t=1}^T$, ..., $\{\tilde{\mathbf{u}}_t^{in}\}_{t=1}^T$, $\{\mathbf{R}_t^{i1}\}_{t=1}^T$, ..., $\{\mathbf{R}_t^{in}\}_{t=1}^T$ $(i=1, ..., n)$, the maximum number of iteration steps for the algorithm, i.e., $k_{max} \geq 1$ and a sufficiently small ε-value that will help to indicate whether or not the algorithm has converged, i.e., has succeeded in approximating a solution of the problem under consideration.

The OPTGAME algorithm approximates the equilibrium solutions of the nonlinear quadratic difference game by following an iterative procedure as frequently used for solving a (one-player) tracking problem (see Fig. 1). For the cooperative Pareto-optimal solution and the noncooperative feedback solutions, this is accomplished by using the method of dynamic programming, which leads to a system of matrix Riccati difference equations which can be rapidly solved. For deriving the open-loop Nash equilibrium solution, Pontryagin's maximum principle is used (see Sect. 3.2).

3.1 The Feedback Nash Equilibrium Solution

The OPTGAME algorithm is initialized with an arbitrary set of tentative paths for all players' control variables,[7] denoted by $\{\bar{\mathbf{u}}_t^1(k)\}_{t=1}^T, ..., \{\bar{\mathbf{u}}_t^n(k)\}_{t=1}^T$ for $k=0$. Then, step by step (i.e., for $k = 1, ..., k_{max}$), the algorithm drives this path towards an approximation of the respective equilibrium path. Using a Gauss-Seidel, a Newton-Raphson, a Levenberg-Marquardt, or a trust region solver, the 'initial' tentative state path is computed by solving $\bar{\mathbf{x}}_t(0) - \mathbf{f}(\bar{\mathbf{x}}_{t-1}(0), \bar{\mathbf{x}}_t(0), \bar{\mathbf{u}}_t^1(0), ..., \bar{\mathbf{u}}_t^n(0), \mathbf{z}_t) = \mathbf{0}$ with $\bar{\mathbf{x}}_{t-1}(0)$ given $\forall\, t \in \{1, ..., T\}$. The resulting reference path for iteration $k=1$ is then given by $\{\bar{\mathbf{X}}_t(1)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(1) := [\,\bar{\mathbf{x}}_t(0) \;\; \bar{\mathbf{u}}_t^1(0) \;\; ... \;\; \bar{\mathbf{u}}_t^n(0)]'$.

Step 1. At iteration step k, we follow [32,33,35] and replace the system of autonomous nonlinear difference equations (Eq. 4) with a non-autonomous linearization of Eq. 4 evaluated along the current reference path, $\{\bar{\mathbf{X}}_t(k)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(k) := [\,\bar{\mathbf{x}}_t(k) \;\; \bar{\mathbf{u}}_t^1(k) \;\; ... \;\; \bar{\mathbf{u}}_t^n(k)]'$ for $k=1, ..., k_{max}$.[8] The parameters of the resulting non-autonomous linear system,

$$\mathbf{x}_t(k) = \mathbf{A}_t(k)\mathbf{x}_{t-1}(k) + \sum_{i=1}^{n} \mathbf{B}_t^i(k)\mathbf{u}_t^i(k) + \mathbf{c}_t(k), \qquad \mathbf{x}_0(k) = \bar{\mathbf{x}}_0, \tag{5}$$

[7] For $k=0$, either zero-vectors may be assigned to $\{\bar{\mathbf{u}}_t^1(k)\}_{t=1}^T$, ..., $\{\bar{\mathbf{u}}_t^n(k)\}_{t=1}^T$ or historical data or other external information may be used.

[8] Note that we do not linearize the nonlinear system prior to executing the optimization procedure but rather linearize the system repeatedly during the optimization process along the current reference path $\{\bar{\mathbf{X}}_t(k)\}_{t=1}^T$ (for $k=1, ..., k_{max}$).

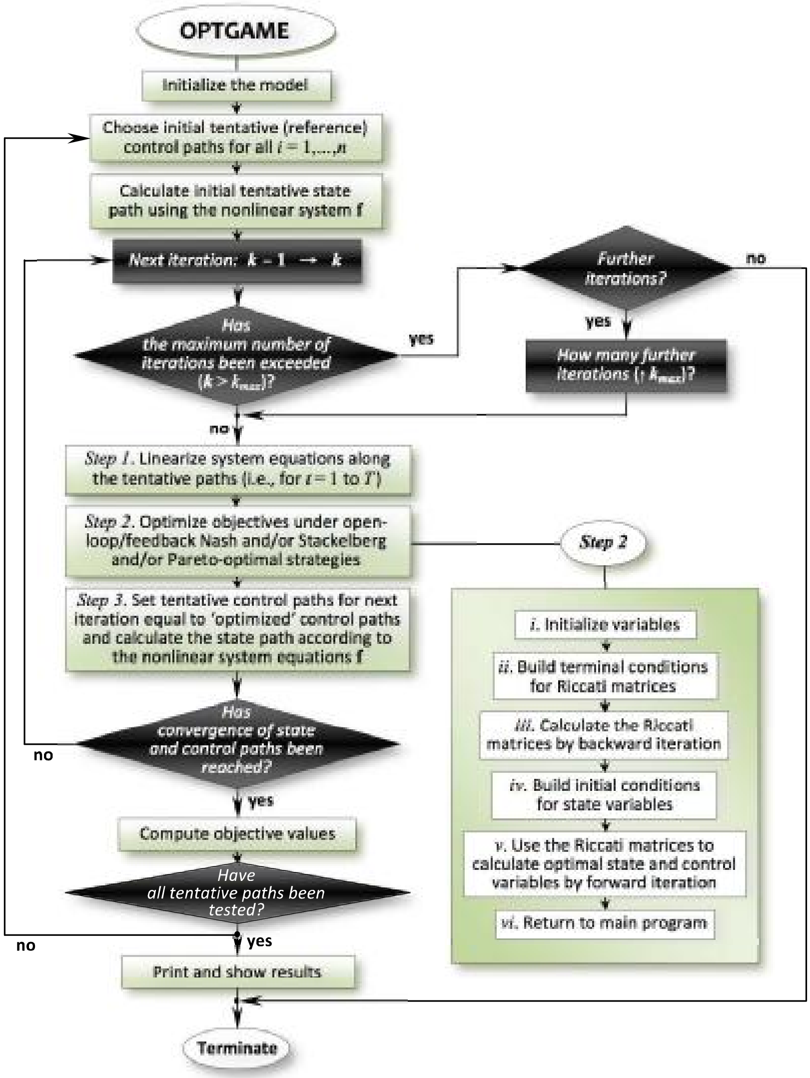

Fig. 1 Flow chart of the OPTGAME algorithm, with k indicating the number of iterations

computed for iteration step k, are time-dependent functions of the reference paths along which they are evaluated, i.e.,

$$\mathbf{A}_t(k) := \left[\mathbf{I} - \mathbf{F}_{\mathbf{x}_t}(\bar{\mathbf{x}}_t(k))\right]^{-1} \mathbf{F}_{\mathbf{x}_{t-1}}(\bar{\mathbf{x}}_{t-1}(k)), \tag{6}$$

$$\mathbf{B}_t^i(k) := \left[\mathbf{I} - \mathbf{F}_{\mathbf{x}_t}(\bar{\mathbf{x}}_t(k))\right]^{-1} \mathbf{F}_{\mathbf{u}_t^i}\left(\bar{\mathbf{u}}_t^i(k)\right), \qquad i=1,...,n, \tag{7}$$

$$\mathbf{c}_t(k) := \bar{\mathbf{x}}_t(k) - \mathbf{A}_t(k)\bar{\mathbf{x}}_{t-1}(k) - \sum_{i=1}^{n} \mathbf{B}_t^i(k)\bar{\mathbf{u}}_t^i(k), \tag{8}$$

with $\mathbf{I}$ denoting an $m_x \times m_x$ identity matrix, and (for $\tau = t,\ t-1$) with

$$\mathbf{F}_{\mathbf{x}_\tau}(\bar{\mathbf{x}}_\tau(k)) := \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}_\tau}\right|_{\bar{\mathbf{x}}_\tau(k)} = \left[\left.\frac{\partial f^1}{\partial \mathbf{x}_\tau}\right|_{\bar{\mathbf{x}}_\tau(k)} \cdots \left.\frac{\partial f^{m_x}}{\partial \mathbf{x}_\tau}\right|_{\bar{\mathbf{x}}_\tau(k)}\right]' \in \mathbb{R}^{m_x \times m_x},$$

$$\mathbf{F}_{\mathbf{u}_t^i}\left(\bar{\mathbf{u}}_t^i(k)\right) := \left.\frac{\partial \mathbf{f}}{\partial \mathbf{u}_t^i}\right|_{\bar{\mathbf{u}}_t^i(k)} = \left[\left.\frac{\partial f^1}{\partial \mathbf{u}_t^i}\right|_{\bar{\mathbf{u}}_t^i(k)} \cdots \left.\frac{\partial f^{m_x}}{\partial \mathbf{u}_t^i}\right|_{\bar{\mathbf{u}}_t^i(k)}\right]' \in \mathbb{R}^{m_x \times m_i}.$$

In the remainder of this subsection, we confine ourselves to a discussion of the *feedback Nash equilibrium solution* and add a description of what has to be changed in order to arrive at the feedback Stackelberg, the open-loop Nash, or the Pareto-optimal solutions in Sect. 3.2.

Step 2. To derive the iteration k values for state and control paths, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$ and $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$ $(i=1,...,n)$ respectively, the algorithm calculates the values of the state and control paths that minimize Eq. 1 subject to Eq. 5 evaluated along the current tentative paths, $\{\bar{\mathbf{x}}_t(k)\}_{t=1}^T$ and $\{\bar{\mathbf{u}}_t^i(k)\}_{t=1}^T$ $(i=1,...,n)$. (For the derivation of the following procedure see Appendix A.1).

Starting with their terminal values, $\mathbf{P}_T^i(k) = \mathbf{Q}_T^i$ and $\mathbf{p}_T^i(k) = \mathbf{Q}_T^i\tilde{\mathbf{x}}_T^i$, the algorithm computes all players' Riccati matrices, $\mathbf{P}_t^i(k)$ and $\mathbf{p}_t^i(k)$ $\forall\, i \in \{1,...,n\}$, which are the parameters of the quadratic value function and which are recursively determined by

$$\mathbf{P}_{t-1}^i(k) = \mathbf{Q}_{t-1}^i + \mathbf{K}_t'(k)\mathbf{P}_t^i(k)\mathbf{K}_t(k) + \sum_{j=1}^{n} \mathbf{G}_t^{j'}(k)\mathbf{R}_t^{ij}\mathbf{G}_t^j(k), \tag{9}$$

$$\mathbf{p}_{t-1}^i(k) = \mathbf{Q}_{t-1}^i\tilde{\mathbf{x}}_{t-1}^i - \mathbf{K}_t'(k)[\mathbf{P}_t^i(k)\mathbf{k}_t(k) - \mathbf{p}_t^i(k)] + \sum_{j=1}^{n} \mathbf{G}_t^{j'}(k)\mathbf{R}_t^{ij}[\tilde{\mathbf{u}}_t^{ij} - \mathbf{g}_t^j(k)], \tag{10}$$

with

$$\mathbf{K}_t(k) := \mathbf{A}_t(k) + \sum_{j=1}^{n} \mathbf{B}_t^j(k)\mathbf{G}_t^j(k), \tag{11}$$

$$\mathbf{k}_t(k) := \mathbf{c}_t(k) + \sum_{j=1}^{n} \mathbf{B}_t^j(k)\mathbf{g}_t^j(k). \tag{12}$$

The feedback matrices $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$ $\forall\, i \in \{1,...,n\}$ are determined by solving the following system of $2n$ linear matrix equations:

$$\mathbf{D}_t^i(k)\mathbf{G}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k)[\mathbf{A}_t(k) + \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbf{B}_t^{j'}(k)\mathbf{G}_t^{j'}(k)] = \mathbf{0}, \tag{13}$$

$$\mathbf{D}_t^i(k)\mathbf{g}_t^i(k) + \mathbf{a}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k) \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbf{B}_t^{j'}(k)\mathbf{g}_t^{j'}(k) = \mathbf{0}, \tag{14}$$

with $\mathbf{A}_t(k)$, $\mathbf{B}_t^i(k)$, and $\mathbf{c}_t(k)$ defined by Eqs. 6, 7, and 8, and with

$$\mathbf{D}_t^i(k) := \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k)\mathbf{B}_t^i(k) + \mathbf{R}_t^{ii}, \tag{15}$$

$$\mathbf{a}_t^i(k) := \mathbf{B}_t^{i'}(k)[\mathbf{P}_t^i(k)\mathbf{c}_t(k) - \mathbf{p}_t^i(k)] - \mathbf{R}_t^{ii}\tilde{\mathbf{u}}_t^{ii}. \tag{16}$$

Using both Riccati matrices, $\mathbf{P}_t^i(k)$ and $\mathbf{p}_t^i(k)$, and feedback matrices, $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$, computed for all players and for all time periods, i.e. $\forall\, i \in \{1,...,n\}$ and $\forall\, t \in \{1,...,T\}$, the k^{th} iteration of the feedback Nash equilibrium path for the state variable, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$, and the k^{th} iteration of player i's equilibrium path for his/her own control variable, $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$, are determined by

$$\hat{\mathbf{x}}_t^*(k) = \mathbf{K}_t(k)\hat{\mathbf{x}}_{t-1}^*(k) + \mathbf{k}_t(k), \tag{17}$$

$$\hat{\mathbf{u}}_t^{i*}(k) = \mathbf{G}_t^i(k)\hat{\mathbf{x}}_{t-1}^*(k) + \mathbf{g}_t^i(k), \tag{18}$$

starting with $\hat{\mathbf{x}}_0^*(k) = \bar{\mathbf{x}}_0$, where $\mathbf{K}_t(k)$ and $\mathbf{k}_t(k)$ are defined by Eq. 11 and Eq. 12 respectively.

Step 3. As long as $k \le k_{max}$, the control paths obtained for iteration k in the previous step (cf. Eq. 18) are used as reference control paths for iteration $k+1$, i.e., $\bar{\mathbf{u}}_t^i(k+1) := \hat{\mathbf{u}}_t^{i*}(k)$ $\forall\ i \in \{1,...,n\}$ $\forall\ t \in \{1,...,T\}$. Then the reference state path for iteration $k+1$ is computed by numerically solving $\bar{\mathbf{x}}_t(k+1) - \mathbf{f}\left(\bar{\mathbf{x}}_{t-1}(k+1), \bar{\mathbf{x}}_t(k+1), \bar{\mathbf{u}}_t^1(k+1), ..., \bar{\mathbf{u}}_t^n(k+1), \mathbf{z}_t\right) = \mathbf{0}$, with respect to the variable $\bar{\mathbf{x}}_t(k+1)$ (where $\bar{\mathbf{x}}_0(k+1) = \bar{\mathbf{x}}_0$).[9] The tentative path along which the $(k+1)^{th}$ linearization procedure will be carried out is given by $\{\bar{\mathbf{X}}_t(k+1)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(k+1) := [\,\bar{\mathbf{x}}_t(k+1)\ \ \hat{\mathbf{u}}_t^{1*}(k)\ \ ...\ \ \hat{\mathbf{u}}_t^{n*}(k)]'$ (cf. Fig. 1, *Step 3*).

Then *Steps 1* and *2* are repeated, and the reference control paths for the $(k+2)^{th}$ iteration are determined by the iteration $k+1$ control paths minimizing Eq. 1 subject to Eq. 5. This yields $\bar{\mathbf{u}}_t^i(k+2) := \hat{\mathbf{u}}_t^{i*}(k+1)$ $\forall\, i \in \{1,...,n\}$ $\forall\, t \in \{1,...,T\}$, and so on.

Let $\exists\ k^* \le k_{max}$, for which the reference path $\{\bar{\mathbf{X}}_t(k^*)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(k^*) := [\bar{\mathbf{x}}_t(k^*)\ \ \bar{\mathbf{u}}_t^1(k^*)\ \ ...\ \ \bar{\mathbf{u}}_t^n(k^*)]' = [\bar{\mathbf{x}}_t(k^*)\ \ \hat{\mathbf{u}}_t^{1*}(k^*-1)\ \ ...\ \ \hat{\mathbf{u}}_t^{n*}(k^*-1)]'$, and $\{\hat{\mathbf{X}}_t^*(k^*-1)\}_{t=1}^T$, computed according to Eqs. 17 and 18, fall into an ε-tube around

[9] Again a Gauss-Seidel, a Newton-Raphson, a Levenberg-Marquardt, or a trust region solver is applied for determining the state variable.

$\{\bar{\mathbf{X}}_t(k^*-1)\}_{t=1}^T$. Then, for these consecutive paths, no variable differs by more than a value of ε. In other words, for each player i, $\bar{\mathbf{u}}_t^i(k^*) \pm \delta_t^i = \bar{\mathbf{u}}_t^i(k^*-1)$ with $\delta_t^i \in \mathbb{R}^{m_i}$, $\delta_t^{ij} \leq \varepsilon$ $(j=1,...,m_i)$, $\forall\ t$. We know, however, that the tentative control paths of iteration k^* are identical to $\{\hat{\mathbf{u}}_t^{i*}(k^*-1)\}_{t=1}^T$. Because of $\hat{\mathbf{x}}_t^*(k^*-1) = \bar{\mathbf{x}}_t(k^*) \pm \delta_t^x$ with $\delta_t^x \in \mathbb{R}^{m_x}$, $\delta_t^{xj} \leq \varepsilon$ $(j = 1, ..., m_x)$ $\forall\ t \in \{1, ..., T\}$, we can infer from this that the state path calculated from the nonlinear system dynamics $\mathbf{f}$ (Eq. 4) using $\{\bar{\mathbf{u}}_t^1(k^*)\}_{t=1}^T, ..., \{\bar{\mathbf{u}}_t^n(k^*)\}_{t=1}^T$ equals the optimized (equilibrium) state path calculated according to Eq. 17 using the linearized system representation (Eq. 5) evaluated along $\{\bar{\mathbf{X}}_t(k^*-1)\}_{t=1}^T$. In this case, $\mathbf{x}_t^* := \hat{\mathbf{x}}_t^*(k^*-1)$ and $\mathbf{u}_t^{i*} := \hat{\mathbf{u}}_t^{i*}(k^*-1)$ $\forall\ i \in \{1, ..., n\}$ $\forall\ t \in \{1, ..., T\}$. In other words, the algorithm has converged,[10] and the paths obtained do indeed minimize Eq. 1 subject to Eq. 4. Then it is reasonable to compute the values of the players' objective functions (evaluated along their respective optimal path, i.e., $\{\mathbf{X}_t^*\}_{t=1}^T$ with $\mathbf{X}_t^* := [\,\mathbf{x}_t^* \ \ \mathbf{u}_t^{1*} \ \ ... \ \ \mathbf{u}_t^{n*}]'$) (cf. Fig. 1).[11]

3.2 Feedback Stackelberg Equilibrium, Open-Loop Nash Equilibrium, and Pareto-Optimal Solutions

To derive the iteration k values of the state and control paths, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$ and $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$ (for $i = 1, ..., n$) respectively, for the feedback Stackelberg solution concept, we utilize the procedure described in Sect. 3.1 for deriving the feedback Nash equilibrium solution but replace *Step 2* by the following procedure.

Feedback Stackelberg Equilibrium Solution. Starting with the terminal conditions, $\mathbf{P}_T^i(k) = \mathbf{Q}_T^i$ and $\mathbf{p}_T^i(k) = \mathbf{Q}_T^i \tilde{\mathbf{x}}_T^i$, we derive the matrices $\mathbf{P}_t^i(k)$ and $\mathbf{p}_t^i(k)$ $\forall\ i \in \{1, ..., n\}$ by backward iteration in t according to Eqs. 9 and 10 respectively, augmented with feedback matrices defined by

$$\mathbf{G}_t^1(k) := \ -[\overline{\mathbf{C}}_t(k)]^{-1}[\overline{\mathbf{B}}_t'(k)\mathbf{P}_t^1(k)\mathbf{A}_t(k) + \sum_{j=2}^n \overline{\mathbf{D}}_t^j(k)\mathbf{W}_t^j(k)], \tag{19}$$

$$\mathbf{g}_t^1(k) := \ -[\overline{\mathbf{C}}_t(k)]^{-1}[\overline{\mathbf{a}}_t(k) + \mathbf{a}_t^1(k) + \sum_{j=2}^n \overline{\mathbf{D}}_t^j(k)\mathbf{w}_t^j(k)], \tag{20}$$

$$\mathbf{G}_t^i(k) := \mathbf{W}_t^i(k) + \mathbf{\Phi}_t^i(k)\mathbf{G}_t^1(k), \qquad i = 2, ..., n, \tag{21}$$

$$\mathbf{g}_t^i(k) := \mathbf{w}_t^i(k) + \mathbf{\Phi}_t^i(k)\mathbf{g}_t^1(k), \qquad i = 2, ..., n, \tag{22}$$

[10] If convergence has not been obtained before k reaches its maximum value, k_{max}, the iterative optimization procedure terminates without succeeding in finding an equilibrium feedback solution. It can then be re-started with an alternative initial tentative control path (to be specified by the user; see Fig. 1).

[11] Under our assumptions, for the linear time-varying dynamic system approximating the nonlinear system, there always exists a (not necessarily unique) feedback Nash equilibrium solution. For the nonlinear dynamic system, this is not always guaranteed, even if the algorithm converges fairly quickly. In any case, alternative tentative initial control paths can be used to provide more insight into the nature of the solution obtained.

with

$$\overline{\mathbf{B}}_t(k) := \mathbf{B}_t^1(k) + \sum_{j=2}^{n} \mathbf{B}_t^j(k)\mathbf{\Phi}_t^j(k), \tag{23}$$

$$\overline{\mathbf{C}}_t(k) := \overline{\mathbf{B}}_t'(k)\mathbf{P}_t^1(k)\overline{\mathbf{B}}_t(k) + \mathbf{R}_t^{11} + \sum_{j=2}^{n} \mathbf{\Phi}_t^{j'}(k)\mathbf{R}_t^{1j}\mathbf{\Phi}_t^j(k), \tag{24}$$

$$\overline{\mathbf{D}}_t^i(k) := \overline{\mathbf{B}}_t'(k)\mathbf{P}_t^1(k)\mathbf{B}_t^i(k) + \mathbf{\Phi}_t^{i'}(k)\mathbf{R}_t^{1i}, \qquad i=2,...,n, \tag{25}$$

$$\overline{\mathbf{a}}_t(k) := \sum_{j=2}^{n} \mathbf{\Phi}_t^{j'}(k)[\mathbf{B}_t^{j'}(k)[\mathbf{P}_t^1(k)\mathbf{c}_t(k) - \mathbf{p}_t^1(k)] - \mathbf{R}_t^{1j}\tilde{\mathbf{u}}_t^{1j}], \tag{26}$$

and $\mathbf{A}_t(k)$, $\mathbf{B}_t^i(k)$, $\mathbf{c}_t(k)$ and $\mathbf{a}_t^1(k)$ given by Eqs. 6, 7, 8, and 16 (for $i = 1$) respectively.[12] The reaction coefficients, $\mathbf{\Phi}_t^i(k) := \partial \mathbf{u}_t^i(k)/\partial \mathbf{u}_t^1(k)$ $(i = 2, ..., n)$ in Eqs. 21−26, are determined as solutions of a system of $n-1$ linear matrix equations,

$$\mathbf{D}_t^i(k)\mathbf{\Phi}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k)[\mathbf{B}_t^1(k) + \sum_{\substack{j=2\\ j\neq i}}^{n} \mathbf{B}_t^j(k)\mathbf{\Phi}_t^j(k)] = \mathbf{0}, \tag{27}$$

where $\mathbf{B}_t^i(k)$ and $\mathbf{D}_t^i(k)$ are defined by Eqs. 7 and 15 respectively.

The matrices $\mathbf{W}_t^2(k), ..., \mathbf{W}_t^n(k)$ and $\mathbf{w}_t^2(k), ..., \mathbf{w}_t^n(k)$ required for computing the feedback matrices given by Eqs. 19−22 are determined as solutions of the following system of $2(n-1)$ linear matrix equations:

$$\mathbf{D}_t^i(k)\mathbf{W}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k)[\mathbf{A}_t(k) + \sum_{\substack{j=2\\ j\neq i}}^{n} \mathbf{B}_t^j(k)\mathbf{W}_t^j(k)], \tag{28}$$

$$\mathbf{D}_t^i(k)\mathbf{w}_t^i(k) + \mathbf{a}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t^i(k)\sum_{\substack{j=2\\ j\neq i}}^{n} \mathbf{B}_t^j(k)\mathbf{w}_t^j(k), \tag{29}$$

with $\mathbf{A}_t(k)$, $\mathbf{B}_t^i(k)$, $\mathbf{D}_t^i(k)$, and $\mathbf{a}_t^i(k)$ determined by Eqs. 6, 7, 15, and 16.

With the Riccati matrices computed for all players and all time periods (cf. Eqs. 9 and 10), and the feedback matrices, $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$ $\forall\ i \in \{1, ..., n\}$ (Eqs. 19−22) for all time periods, the feedback Stackelberg equilibrium values for the state and the control variables can be determined by utilizing the functional forms given by Eqs. 17 and 18 respectively. If the algorithm has converged (cf. Sect. 3.1), for $i = 1, ..., n$, the values of the objective functions are calculated according to Eq. 1 (evaluated along their respective optimal path). This allows, for example, the identification of the size of the 'leader advantage' of a dominant player by comparing the values of the objective function (Eq. 1) evaluated along the Nash and Stackelberg equilibrium paths.

[12] As by assumption (cf. Sect. 2.1) the leader's penalty matrix $\mathbf{R}_t^{11}$ is required to be of full rank, the $m_1 \times m_1$ matrix $\overline{\mathbf{C}}_t(k)$ (see Eq. 24) exists and is invertible.

A brief outline of what has to be changed in *Step 2* to utilize the procedure described in Sect. 3.1 for obtaining approximations of the state and control paths in the open-loop Nash equilibrium is given next.[13] Again, at iteration step k, we linearize the nonlinear autonomous system $\mathbf{f}$ given by Eq. 4, along the tentative paths for state and controls, jointly embodied in $\{\bar{\mathbf{X}}_t(k)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(k) := [\bar{\mathbf{x}}_t(k) \;\; \bar{\mathbf{u}}_t^1(k) \;\; ... \;\; \bar{\mathbf{u}}_t^n(k)]'$.

Open-Loop Nash Equilibrium Solution. Let the matrix

$$\mathbf{C}_t(k) := \mathbf{I} + \sum_{j=1}^{n} \mathbf{B}_t^j(k)[\mathbf{R}_t^{jj}]^{-1}\mathbf{B}_t^{j'}(k)\mathbf{P}_t^j(k) \tag{30}$$

exist and be invertible, with $\mathbf{I}$ denoting an $m_x \times m_x$ identity matrix. Starting with the conditions, $\mathbf{P}_T^i(k) = \mathbf{Q}_T^i$ and $\mathbf{p}_T^i(k) = -\mathbf{Q}_T^i\tilde{\mathbf{x}}_T^i$, all players' Riccati matrices are computed backwards in time according to a system of recursive matrix equations,

$$\mathbf{P}_{t-1}^i(k) = \mathbf{Q}_{t-1}^i + \mathbf{A}_t'(k)\mathbf{P}_t^i(k)[\mathbf{C}_t(k)]^{-1}\mathbf{A}_t(k), \tag{31}$$

$$\mathbf{p}_{t-1}^i(k) = -\mathbf{Q}_{t-1}^i\tilde{\mathbf{x}}_{t-1}^i + \mathbf{A}_t'(k)[\mathbf{P}_t^i(k)[\mathbf{C}_t(k)]^{-1}\mathbf{b}_t(k) + \mathbf{p}_t^i(k)], \tag{32}$$

with $\mathbf{A}_t$, $\mathbf{B}_t^i$, and $\mathbf{c}_t$ given by Eqs. 6, 7, and 8, and with

$$\mathbf{b}_t(k) := \mathbf{c}_t(k) + \sum_{j=1}^{n} \mathbf{B}_t^j(k)[\tilde{\mathbf{u}}_t^{jj} - [\mathbf{R}_t^{jj}]^{-1}\mathbf{B}_t^{j'}(k)\mathbf{p}_t^j(k)]. \tag{33}$$

With these matrices $\mathbf{P}_t^1(k), ..., \mathbf{P}_t^n(k), \mathbf{p}_t^1(k), ..., \mathbf{p}_t^n(k)$ computed for all players and for all time periods, i.e. $\forall\, i = 1, ..., n$ and $\forall\, t = 1, ..., T$, the state variable is determined by

$$\hat{\mathbf{x}}_t^*(k) = [\mathbf{C}_t(k)]^{-1}[\mathbf{A}_t(k)\hat{\mathbf{x}}_{t-1}^*(k) + \mathbf{b}_t(k)], \tag{34}$$

starting with $\hat{\mathbf{x}}_0^*(k) = \bar{\mathbf{x}}_0$. At the beginning of the open-loop Nash game, each of the n simultaneously acting players makes a binding commitment to stick, for the entire planning horizon, to the policy rule

$$\hat{\mathbf{u}}_t^{i*}(k) = \tilde{\mathbf{u}}_t^{ii} - [\mathbf{R}_t^{ii}]^{-1}\mathbf{B}_t^{i'}(k)[\mathbf{P}_t^i(k)\hat{\mathbf{x}}_t^*(k) + \mathbf{p}_t^i(k)], \tag{35}$$

starting with $\hat{\mathbf{x}}_1^*(k) = [\mathbf{C}_1(k)]^{-1}[\mathbf{A}_1(k)\bar{\mathbf{x}}_0 + \mathbf{b}_1(k)]$. These control variables obtained for the k^{th} iteration are used to define the tentative control variables for the $(k+1)^{th}$ iteration, i.e., $\{\bar{\mathbf{u}}_t^i(k+1)\}_{t=1}^T := \{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$, while $\{\bar{\mathbf{x}}_t(k+1)\}_{t=1}^T$ is computed by solving the nonlinear system (cf. Eq. 4). Thus, the tentative path along which the $(k+1)^{th}$ linearization procedure is carried out is given by $\{\bar{\mathbf{X}}_t(k+1)\}_{t=1}^T$ with $\bar{\mathbf{X}}_t(k+1) := [\bar{\mathbf{x}}_t(k+1) \;\; \bar{\mathbf{u}}_t^1(k+1) \;\; ... \;\; \bar{\mathbf{u}}_t^n(k+1)]'$ (cf. Fig. 1). When

[13] Note that the open-loop Nash equilibrium solution of the linearized quadratic game is determined using Pontryagin's maximum principle, while the feedback Nash and feedback Stackelberg solutions of the linearized quadratic game are approximated using the dynamic programming technique.

the algorithm has converged (see Sect. 3.1), the resulting open-loop Nash equilibrium solution has the property that none of the players can reduce his/her individual loss by one-sided deviations from that path. A comparison of the objective function values (Eq. 1) for the open-loop and the feedback Nash equilibrium solutions can yield information about gains and losses associated with the commitment to a certain strategy over several periods of time. Such calculations for different scenarios may also help to shed light on a particular player's incentive for reneging.

The last solution concept for which *Step 2* is discussed here is the cooperative Pareto-optimal solution, which is actually an optimal control rather than a game-theoretic problem. We have, however, decided to include it since it provides valuable information when used as a benchmark.

Again we follow the procedure described in Sect. 3.1 but replace *Step 2* by the following routine where the approximate solution of the Pareto-optimal game (with n cooperating players) is determined by solving a classical optimal control problem using dynamic programming:

Pareto-Optimal Solution. Starting with $\mathbf{P}_T(k) = \sum_{i=1}^{n} \mu^i \mathbf{Q}_T^i$ and $\mathbf{p}_T(k) = \sum_{i=1}^{n} \mu^i \mathbf{Q}_T^i \tilde{\mathbf{x}}_T^i$ respectively, the Riccati matrices of the tracking problem (formulated in terms of Eq. 3 subject to Eq. 4) are computed for all players by iterating backwards in time according to the following system of $2n$ recursive matrix equations,

$$\mathbf{P}_{t-1}(k) = \sum_{i=1}^{n} \mu^i \mathbf{Q}_{t-1}^i + \mathbf{K}_t'(k)\mathbf{P}_t(k)\mathbf{K}_t(k) + \sum_{i=1}^{n} \mathbf{G}_t^{i'}(k)\mathbf{R}_t^i \mathbf{G}_t^i(k), \tag{36}$$

$$\mathbf{p}_{t-1}(k) = \sum_{i=1}^{n} \mu^i \mathbf{Q}_{t-1}^i \tilde{\mathbf{x}}_{t-1}^i + \mathbf{K}_t'(k)[\mathbf{P}_t(k)\mathbf{k}_t(k) - \mathbf{p}_t(k)] + \sum_{i=1}^{n} \mathbf{G}_t^{i'}(k)[\mathbf{r}_t^i - \mathbf{R}_t^i \mathbf{g}_t^i(k)], \tag{37}$$

with $\mathbf{R}_t^i := \sum_{j=1}^{n} \mu^j \mathbf{R}_t^{ji}$, $\mathbf{r}_t^i := \sum_{j=1}^{n} \mu^j \mathbf{R}_t^{ji} \tilde{\mathbf{u}}_t^{ji}$, and $\mathbf{K}_t(k)$ and $\mathbf{k}_t(k)$ defined by Eqs. 11 and 12 respectively. For all $t \in \{1, ..., T\}$, the feedback matrices $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$ $\forall\, i \in \{1, ..., n\}$ in Eqs. 36−37 are determined as solutions of a system of $2n$ linear matrix equations,

$$\overline{\overline{\mathbf{D}}}_t^i(k)\mathbf{G}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t(k)[\mathbf{A}_t(k) + \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{B}_t^{j'}(k)\mathbf{G}_t^{j'}(k)] = \mathbf{0}, \tag{38}$$

$$\overline{\overline{\mathbf{D}}}_t^i(k)\mathbf{g}_t^i(k) + \overline{\overline{\mathbf{a}}}_t^i(k) + \mathbf{B}_t^{i'}(k)\mathbf{P}_t(k) \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{B}_t^{j'}(k)\mathbf{g}_t^{j'}(k) = \mathbf{0}, \tag{39}$$

with

$$\overline{\overline{\mathbf{D}}}_t^i(k) := \mathbf{B}_t^{i'}(k)\mathbf{P}_t(k)\mathbf{B}_t^i(k) + \sum_{j=1}^{n} \mu^j \mathbf{R}_t^{ji}, \tag{40}$$

$$\bar{\bar{\mathbf{a}}}_t^i(k) := \mathbf{B}_t^{i'}(k)[\mathbf{P}_t(k)\mathbf{c}_t(k) - \mathbf{p}_t(k)] - \sum_{j=1}^{n} \mu^j \mathbf{R}_t^{ji} \tilde{\mathbf{u}}_t^{ji}, \tag{41}$$

and $\mathbf{A}_t(k)$, $\mathbf{B}_t^i(k)$, $\mathbf{c}_t(k)$ given by Eqs. 6, 7, and 8 respectively.

With both Riccati matrices $\mathbf{P}_t^i(k)$ and $\mathbf{p}_t^i(k)$ and feedback matrices $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$ computed for all players and for all time periods, i.e. $\forall\ i \in \{1, ..., n\}$ and $\forall\ t \in \{1, ..., T\}$, the k^{th} iteration of the Pareto-optimal path of the state variable, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$, and the k^{th} iteration of player i's Pareto-optimal path of the control variables, $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$, are determined using Eqs. 17 and 18 respectively.

The possibilities and the limitations of the OPTGAME algorithm are similar to the ones observed for the tracking problem in regulator theory. Only if the algorithm converges is the result obtained an approximation of the solution of the original problem [33], and, as Stephen Boyd remarks in his course on Linear Dynamical Systems, the algorithm '*sometimes converges, sometimes to globally optimal control*' ([36], p. 31). We have used the OPTGAME2 algorithm in various experiments with small numerical economic models and obtained convergence in nearly all cases. The results are generally reasonable and can be well interpreted in terms of economic policy, so we are confident that it can also deliver useful insights into more sophisticated policy problems.

4 An Example

In this section we present an application of the OPTGAME algorithm to a monetary union macroeconomic model. Dynamic games have been used by several authors (e.g. [11,31]) to model conflicts between monetary and fiscal policies. There is also a large body of literature on dynamic conflicts between policy makers from different countries on issues of international stabilization (e.g. [12,10,37]). Both types of conflict are present in a monetary union, because a supranational central bank interacts strategically with sovereign governments as national fiscal policy makers in the member states. We use a small stylized nonlinear two-country macroeconomic model of a monetary union (called MUMOD1) to analyse the interactions between fiscal (governments) and monetary (common central bank) policy makers, assuming different objective functions of these decision makers. Using the OPTGAME algorithm we calculate approximate solutions for the four game strategies available in OPTGAME, viz. the cooperative Pareto optimal solution, the open-loop and feedback Nash equilibrium solution and the feedback Stackelberg equilibrium solution. Applying the OPTGAME algorithm to the MUMOD1 model we can show how the policy makers react optimally to demand and supply shocks. For details and additional policy experiments, see [38].

4.1 The MUMOD1 model

In the following brief model description, capital letters indicate nominal values, while lower case letters correspond to real values. Variables are denoted by Roman letters and model parameters are denoted by Greek letters. Three active policy makers are considered: the governments of the two countries responsible for decisions about fiscal policy and the common central bank of the monetary union controlling monetary policy. The two countries are labeled 1 and 2 or core and periphery respectively. Thus MUMOD1 is a stylized model of a monetary union consisting of two homogeneous blocs of countries, which in the current European context might be identified with a stability-oriented bloc (core) and a bloc of countries with problems due to high public debt.

Real output (or the deviation of short-run output from a long-run growth path) in country i $(i = 1, 2)$ at time t $(t = 1, ..., T)$ is determined by a reduced form demand-side equation

$$y_{it} = \delta_i(\pi_{jt} - \pi_{it}) - \gamma_i(r_{it} - \theta) + \rho_i y_{jt} - \beta_i \pi_{it} + \kappa_i y_{i(t-1)} - \eta_i g_{it} + zd_{it}, \quad (42)$$

for $i, j = 1, 2$; $i \neq j$. The variable π_{it} denotes the rate of inflation in country i, r_{it} represents country i's real rate of interest and g_{it} denotes country i's real fiscal surplus (or, if negative, its fiscal deficit) measured in relation to real GDP. g_{it} is assumed to be country i's fiscal policy instrument or control variable. The natural real rate of output growth, $\theta \in [0, 1]$, is assumed to be equal to the natural real rate of interest. The parameters $\delta_i, \gamma_i, \rho_i, \beta_i, \kappa_i, \eta_i$ are assumed to be positive. The variables zd_{1t} and zd_{2t} are non-controlled exogenous variables and represent demand-side shocks in the goods market.

The current real rate of interest for country i $(i = 1, 2)$ is given by

$$r_{it} = I_{it} - \pi_{it}^e, \quad (43)$$

where π_{it}^e denotes the expected rate of inflation in country i and I_{it} denotes the nominal interest rate for country i, which is given by

$$I_{it} = R_{Et} - \lambda_i g_{it} + \chi_i D_{it}, \quad (44)$$

where R_{Et} denotes the prime rate determined by the central bank of the monetary union (its control variable); $-\lambda_i$ and χ_i (λ_i and χ_i are assumed to be positive) are risk premiums for country i's fiscal deficit and public debt level.

The inflation rates for each country $i = 1, 2$ and $t = 1, ..., T$ are determined according to an expectations-augmented Phillips curve

$$\pi_{it} = \pi_{it}^e + \xi_i y_{it}, \quad (45)$$

where ξ_1 and ξ_2 are positive parameters, π_{it}^e denotes the rate of inflation in country i expected to prevail during time period t, which is formed at (the end of) time period $t-1$. Inflationary expectations are formed according to the hypothesis of adaptive expectations:

$$\pi_{it}^e = \varepsilon_i \pi_{i(t-1)} + (1 - \varepsilon_i)\pi_{i(t-1)}^e, \quad (46)$$

where $\varepsilon_i \in [0,1]$ are positive parameters determining the speed of adjustment of expected to actual inflation.

The average values of output and inflation in the monetary union are given by

$$y_{Et} = \omega y_{1t} + (1-\omega) y_{2t}, \ \ \omega \in [0,1], \tag{47}$$

$$\pi_{Et} = \omega \pi_{1t} + (1-\omega) \pi_{2t}, \ \ \omega \in [0,1]. \tag{48}$$

The parameter ω expresses the weight of country 1 in the economy of the whole monetary union as defined by its output level.

The government budget constraint of country i ($i = 1, 2$) is given as

$$D_{it} = (1 + r_{i(t-1)}) D_{i(t-1)} - g_{it}, \tag{49}$$

where D_i denotes real public debt of country i measured in relation to (real) GDP.

Both national fiscal authorities are assumed to care about stabilizing inflation (π), output (y), debt (D) and fiscal deficits in their own countries (g) at each time t. The common central bank is interested in stabilizing inflation and output in the entire monetary union, also taking into account a goal of low and stable interest rates in the union.

Equations (42)-(49) constitute a dynamic game with three players, each of them having one control variable. The model contains 14 endogenous variables and four exogenous variables and is assumed to be played over a finite time horizon.

The parameters of the model are specified for a slightly asymmetric monetary union, see Table 1. Here an attempt has been made to calibrate the model parameters so as to fit the EMU.

Table 1 Parameter values for an asymmetric monetary union, $i = 1, 2$

T	θ	ω	$\delta_i, \beta_i, \eta_i, \varepsilon_i$	$\gamma_i, \rho_i, \kappa_i, \xi_i, \lambda_i$	χ_i
30	3	0.6	0.5	0.25	0.0125

The initial values of the macroeconomic variables, which are the state variables of the dynamic game model, are presented in Table 2. The desired or ideal values assumed for the objective variables of the players are given in Table 3. Country 1 (the core bloc) has an initial debt level of 60% of GDP and aims to decrease this level in a linear way over time to arrive at a public debt of 50% at the end of the planning horizon. Country 2 (the periphery bloc) has an initial debt level of 80% of GDP and aims to decrease its level to 60% at the end of the planning horizon. The ideal rate of inflation is calibrated at 1.8%, which corresponds to the Eurosystem's aim of keeping inflation below, but close to, 2%. The initial values of the two blocs' government debts correspond to those at the beginning of the Great Recession, the recent financial and economic crisis. Otherwise, the initial situation is assumed to be close to equilibrium, with parameter values calibrated accordingly.

Table 2 Initial values of the two-country monetary union

$y_{i,0}$	$\pi_{i,0}$	$\pi^e_{i,0}$	$D_{1,0}$	$D_{2,0}$
0	2	2	60	80

Table 3 Target values for an asymmetric monetary union

$\tilde{y}_{it}$	$\tilde{D}_{1t}$	$\tilde{D}_{2t}$	$\tilde{\pi}_{it}$	$\tilde{\pi}_{Et}$	$\tilde{y}_{Et}$	$\tilde{g}_{it}$	$\tilde{R}_{Et}$
0	60$\searrow$50	80$\searrow$60	1.8	1.8	0	0	3

4.2 Effects of a negative demand-side shock

The MUMOD1 model can be used to simulate the effects of different shocks acting on the monetary union and the effects of policy reactions towards these shocks. Here we investigate a symmetric shock which occurs on the demand side (zd_i) as given in Table 4. The numbers can best be interpreted as percentage points of real GDP.

Table 4 Negative symmetric shock on the demand side

t	1	2	3	4	5	6	$\cdots$	30
zd_1	-2	-4	-2	0	0	0	$\cdots$	0
zd_2	-2	-4	-2	0	0	0	$\cdots$	0

Using the OPTGAME algorithm, we calculate five different solutions: a baseline solution with the shock but with policy instruments held at pre-shock levels (zero for the fiscal balance, 3 for the central bank's interest rate), three noncooperative game solutions and one cooperative game solution. Figures 2-6 show the simulation and optimization results of this experiment. Figures 2-3 show the results for the control variables of the players and Figures 4-6 show the results of selected state variables: output, inflation, and public debt.

Without policy intervention (baseline scenario, denoted by 'simulation'), both countries suffer from the economic downturn modeled by the demand-side shock in the first periods. The output of both countries drops by more than 6%, which is a fairly good approximation of what happened in reality for several European countries. Even more dramatic is the development of public debt. Without policy intervention it increases during the whole planning horizon and arrives at levels of 240% of GDP for country 1 (or the core bloc) and 390% for country 2 (or the periphery bloc), which shows a need for policy actions to preserve the solvency of the governments in the monetary union.

The reactions of the players (the central bank and the governments of the countries) to the demand-side shocks and their intensity depend on the presence or absence of cooperation. Optimal monetary policy has to be expansionary (lowering the prime rate) in all solution concepts considered, but in the cooperative Pareto solution it is more expansionary during the first 15

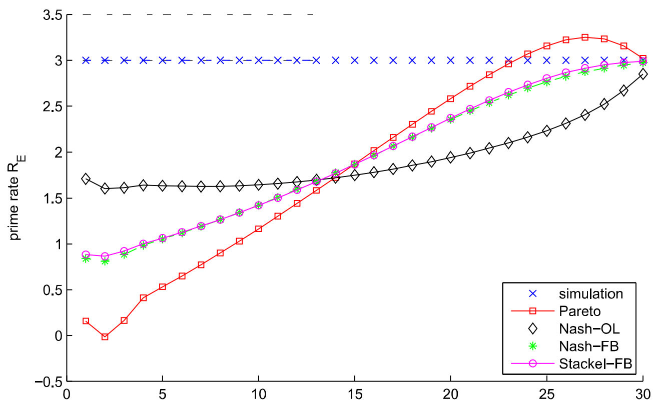

Fig. 2 Prime rate R_{Et} controlled by the central bank

periods. The Nash open-loop equilibrium solution, in contrast, is more or less constant during the whole optimization period, which causes the central bank to be less active at the beginning and relatively more active at the end of the optimization horizon.

With respect to fiscal policy, both countries are required to set expansionary actions and to create deficits in the first three periods in order to absorb the demand-side shock. After that a trade-off between output and public debt occurs and the governments have to take care of the financial situation and produce primary surpluses. The only exception is the cooperative Pareto solution: cooperation between the countries and the central bank (which in this strategy runs a more active expansionary monetary policy) and the resulting moderate inflation means that the balance of public finances can be held close to zero. For country 2 it is even optimal to run a slightly expansionary fiscal policy again during the last 15 periods in the Pareto solution. In spite of this the countries are able to stabilize and bring down their public debts close to the targeted values under cooperation.

In the open-loop Nash equilibrium solution, as the central bank is less active than in all the other solutions, the governments are forced to run restrictive fiscal policies. Here the trade-off between output and the public debt target is dominated by the latter. The lack of cooperation between the players and the open-loop information pattern make the policy makers less flexible and as a result produce large drops in output and an unsustainable deflation. An economic reason for this result is the lack of strong time consistency of strategies in this solution concept, which implies very restrictive fiscal policies.

The noncooperative Nash feedback and Stackelberg feedback solutions give very similar results. In comparison to the Pareto optimal solution, the central bank is less active and the countries run more active fiscal policies (except during the negative demand shock). As a result, output and inflation are

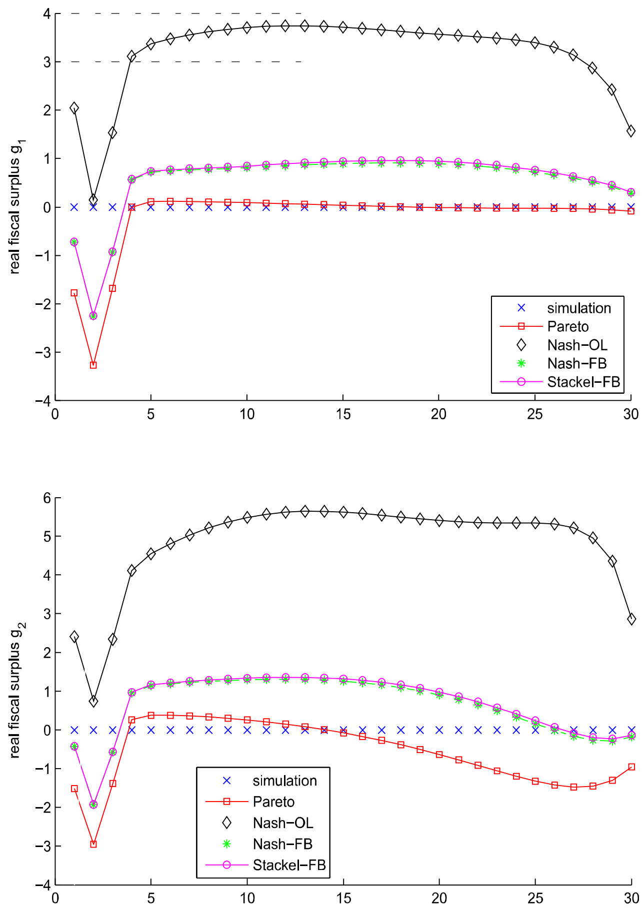

Fig. 3 Country i's fiscal surplus g_{it} (control variable) for $i = 1$ (core; top) and $i = 2$ (periphery; bottom)

slightly below the values achieved in the cooperative solution, and public debt is slightly higher.

Several other policy experiments were run with this model, including supply side shocks and policy reactions to them [38] or various solutions to the emerging public debt crisis such as debt reliefs. As for the OPTGAME algorithm, it can be concluded that it runs smoothly and converged for all policy experiments run. The results concur with economic intuition and although we can, of course, not be sure sure how close the approximate solutions are to the true equilibrium solutions, the conjecture that they are so seems to be justified.

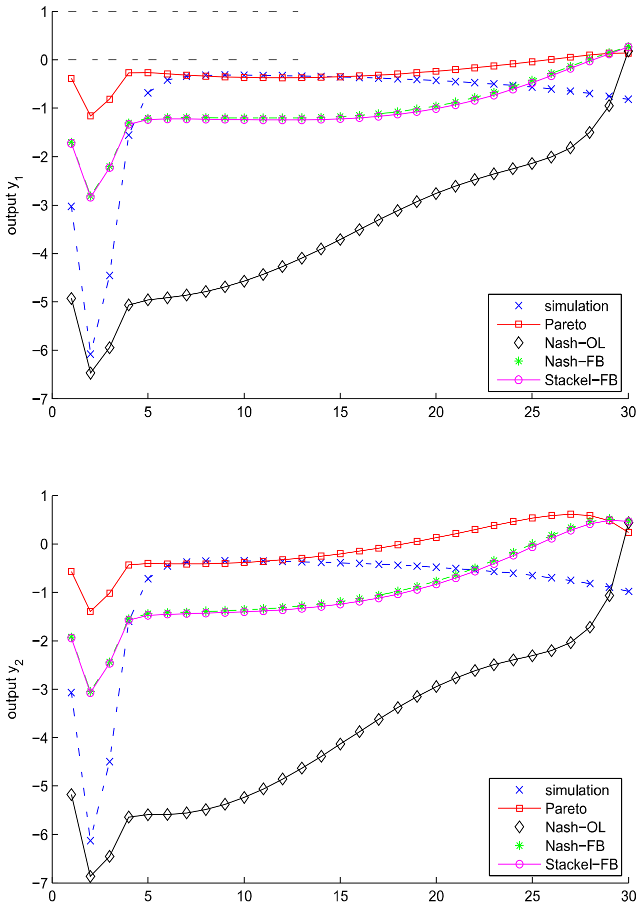

Fig. 4 Country i's output y_{it} for $i = 1$ (core; top) and $i = 2$ (periphery; bottom)

5 Concluding Remarks

In this paper we gave a description of a numerical algorithm which was developed to investigate dynamic game problems for economic policy questions. OPTGAME approximates the solutions of multi player 'tracking games' by iteratively applying a local numerical linearization procedure. The noncooperative solution concepts considered here include Stackelberg and Nash equilibrium solutions, the latter for both open-loop and feedback information patterns. These types of noncooperative games can be used to model the interaction of a finite number of decision makers who share a joint 'environment' but aim at unilaterally minimizing deviations from an individually desirable (multi objective) state or situation over a finite planning horizon. To deter-

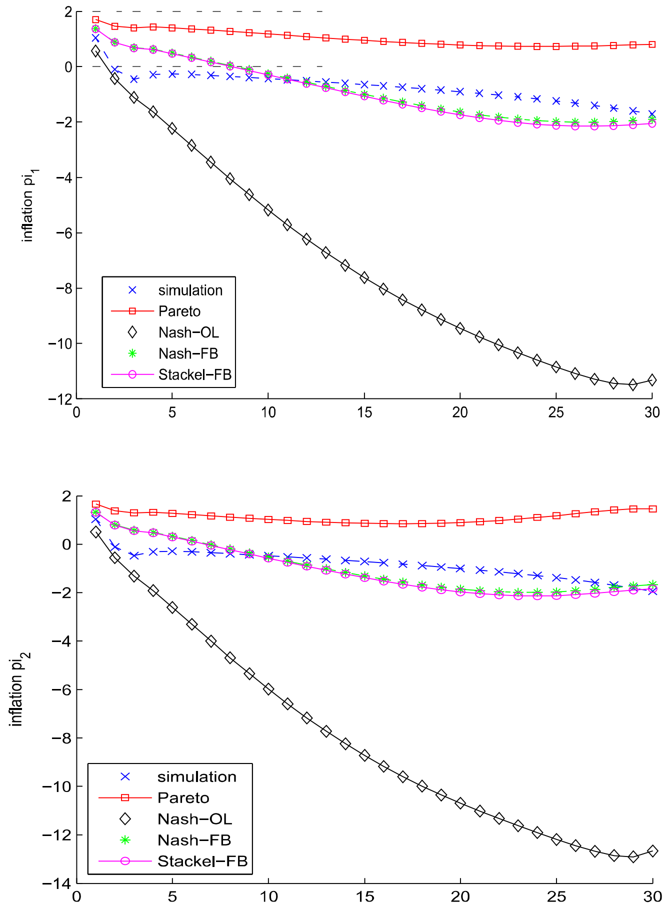

Fig. 5 Country i's inflation rate π_{it} for $i = 1$ (core; top) and $i = 2$ (periphery; bottom)

mine who wins and who loses at whose expense, the Pareto-optimal solution is also included, where the mathematical model that serves as the basis for the iterative numerical procedure for approximating equilibrium solutions is equivalent to the tracking problem in linear quadratic regulator theory.

Several promising extensions of the algorithm are topics for further research, including stochastic dynamic systems and systems with forward-looking (rational) expectations. Moreover, any increase in the number of players increases the number and the possible structure of their interactions, allowing, among others, the formation of hierarchies and coalitions. Extending the OPTGAME algorithm to implement the possibility of a multi-level hierarchical structure and a model of coalition formation constitutes a challenging and important task for future research.

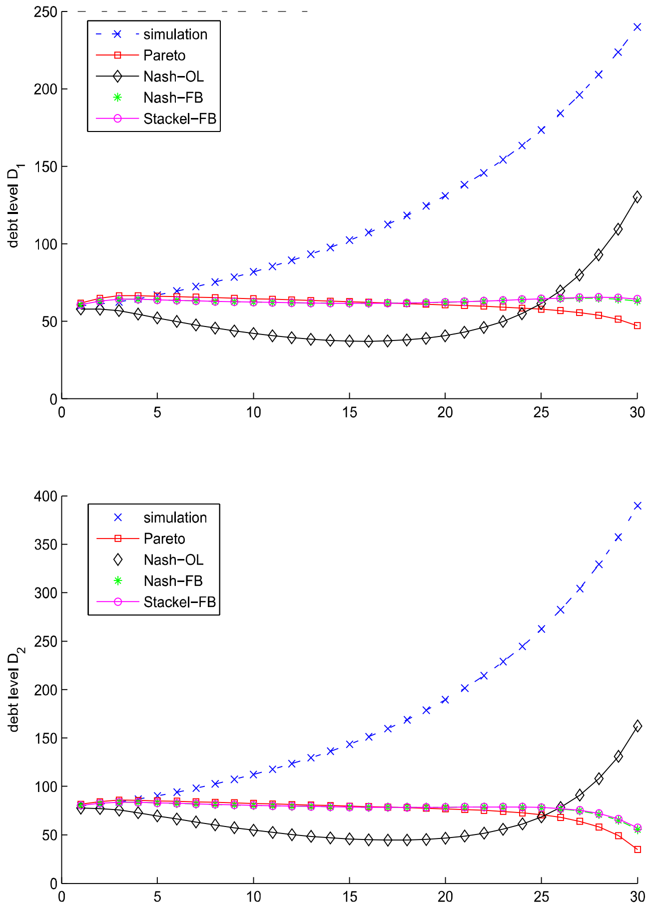

Fig. 6 Country i's debt level D_{it} for $i = 1$ (core; top) and $i = 2$ (periphery; bottom)

References

1. T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. SIAM, Philadelphia, 2 edition, 1999.
2. E. Dockner, S. Jorgensen, N. V. Long, and G. Sorger. *Differential Games in Economics and Management Science*. Cambridge University Press, Cambridge, 2000.
3. U. Doraszelski and M. Satterthwaite. Computable Markov-perfect industry dynamics. *RAND Journal of Economics*, 41(2):215–243, 2010.
4. C. Fershtman and A. Pakes. Dynamic games with asymmetric information: A framework for empirical work. *Quarterly Journal of Economics*, 127(4):1611–1661, 2012.
5. U. Doraszelski and S. Markovich. Advertising dynamics and competitive advantage. *RAND Journal of Economics*, 38(3):557–592, 2007.
6. S. Jørgensen and G. Zaccour. *Differential Games in Marketing*. Kluwer, Boston, 2004.
7. N. Acoccella, G. Di Bartolomeo, and A. Hughes Hallett. *The Theory of Economic Policy in a Strategic Context*. Cambridge University Press, Cambridge, 2013.

8. D. Blueschke and R. Neck. "Core" and "periphery" in a monetary union: A macroeconomic policy game. *International Advances in Economic Research*, 17(3):334–346, 2011.
9. R. Neck and D. A. Behrens. Macroeconomic policies in a monetary union: A dynamic game. *Central European Journal of Operations Research*, 12(2):171–186, 2004.
10. R. Neck and D. A. Behrens. A macroeconomic policy game for a monetary union with adaptive expectations. *Atlantic Economic Journal*, 37(4):335–349, 2009.
11. M. L. Petit. *Control Theory and Dynamic Games in Economic Policy Analysis*. Cambridge University Press, Cambridge, 1990.
12. J. Plasmans, J. Engwerda, B. van Aarle, G. Di Bartolomeo, and T. Michalak. *Dynamic Modeling of Monetary and Fiscal Cooperation Among Nations*. Dynamic Modeling and Econometrics in Economics and Finance, Vol. 8. Springer, Dordrecht, 2006.
13. N. Van Long. Dynamic games in the economics of natural resources: A survey. *Dynamic Games and Applications*, 1(1):115–148, 2011.
14. D. A. Behrens, J. P. Caulkins, G. Feichtinger, and G. Tragler. Incentive Stackelberg strategies for a dynamic game on terrorism. In S. Jørgensen, M. Quincampoix, and V.L. Thomas, editors, *Advances in Dynamic Game Theory*, Annals of the International Society of Dynamic Games, Vol. 9, pages 459–486. Birkhäuser, Boston, 2007.
15. D. Grass, J. P. Caulkins, G. Feichtinger, G. Tragler, and D. A. Behrens. *Optimal Control of Nonlinear Processes. With Applications in Drugs, Corruption, and Terror*. Springer, Berlin/Heidelberg, 2008.
16. A. J. Novak, G. Feichtinger, and G. Leitmann. A differential game related to terrorism: Nash and Stackelberg strategies. *Journal of Optimization Theory and Applications*, 144(3):533–555, 2010.
17. J. C. Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley and Sons, Chichester, 2005.
18. M. Pachter and K. D. Pham. Discrete-time linear-quadratic dynamic games. *Journal of Optimization Theory and Applications*, 146(1):151–179, 2010.
19. W. J. McKibbin. Solving rational expectations models with and without strategic behavior. *Reserve Bank of Australia Research*, Discussion Paper 8706, 1987.
20. T. Michalak, J. Engwerda, and J. Plasmans. A numerical toolbox to solve n-player affine LQ open-loop differential games. *Computational Economics*, 37:375–410, 2011.
21. A. Pakes, G. Gowrisankaran, and P. McGuire. Implementing the Pakes-McGuire algorithm for computing Markov-perfect equilibria in GAUSS. Working Paper, Department of Economics, Yale University, 1993.
22. A. Pakes and P. McGuire. Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND Journal of Economics*, 25:555–589, 1994.
23. A. Pakes and P. McGuire. Stochastic algorithms, symmetric Markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica*, 69:1261–1281, 2001.
24. R. Ericson and A. Pakes. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82, 1995.
25. V. Farias, D. Saure, and G. Y. Weintraub. An approximate dynamic programming approach to solving dynamic oligopoly models. *RAND Journal of Economics*, 43(2):253–282, 2012.
26. B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear-Quadratic Methods*. Prentice Hall, New York, 1998.
27. M. Pachter. Revisit of linear-quadratic optimal control. *Journal of Optimization Theory and Applications*, 140(2):301–314, 2009.
28. A. J. de Zeeuw and F. van der Ploeg. Difference games and policy optimization: A conceptional framework. *Oxford Economic Papers*, 43(4):612–636, 1991.
29. R. E. Lucas. Econometric policy evaluation: A critique. In K. Brunner and A. H. Meltzer, editors, *Theory, Policy, Institutions: Papers from the Carnegie-Rochester Conference Series on Public Policy*, pages 257–284. North-Holland, Amsterdam, 1983.
30. D. A. Behrens, M. Hager, and R. Neck. OPTGAME 1.0: A numerical algorithm to determine solutions for two-person difference games. In R. Neck, editor, *Modelling and Control of Economic Systems 2001. A Proceedings volume from the 10th IFAC Symposium, Klagenfurt, Austria, 6-8 September 2001*, pages 47–58. Elsevier, Oxford, 2003.

31. M. Hager, R. Neck, and D. A. Behrens. Solving dynamic macroeconomic policy games using the algorithm OPTGAME 1.0. *Optimal Control Applications and Methods*, 22(5-6):301 − 332, 2001.
32. G. C. Chow. *Analysis and Control of Dynamic Economic Systems*. John Wiley & Sons, New York, 1975.
33. G. C. Chow. *Econometric Analysis by Control Methods*. John Wiley & Sons, New York, 1981.
34. J. F. Reinganum and N. S. Stokey. Oligopoly extraction of a common property natural resource: The importance of the period of commitment in dynamic games. *International Economic Review*, 26(1):161–173, 1985.
35. J. Matulka and R. Neck. OPTCON: An algorithm for the optimal control of nonlinear stochastic models. *Annals of Operations Research*, 37:375–401, 1992.
36. S. P. Boyd. Linear dynamical systems. Linear quadratic regulator: Discrete-time finite horizon. See website http://www.stanford.edu/class/ee363, 2005.
37. K. Hamada and M. Kawai. International economic policy coordination: Theory and policy implications. In M. U. Fratianni, D. Salvatore, and J. von Hagen, editors, *Macroeconomic Policy in Open Economies*, pages 87–147. Greenwood Press, Westport, CT, 1997.
38. R. Neck and D. Blueschke. Policy interactions in a monetary union: An application of the OPTGAME algorithm. In J. Haunschmied, V. M. Veliov, and S. Wrzaczek, editors, *Dynamic Games in Economics*. Springer, Berlin, forthcoming 2013.

A Appendix

A.1 Derivation of the Feedback Nash Equilibirum Solution for a LQDG

All players have access to the complete state information and seek control rules that respond to the currently observed state. Here we will describe the corresponding feedback Nash equilibrium solution for iteration step k, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$ and $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$ $\forall\ i \in \{1,...,n\}$, that minimizes Eq. 1 subject to Eq. 5 by applying the method of dynamic programming. We set up player i's $(i=1...,n)$ individual cost-to-go function for the terminal period, T,

$$J_T^i(k) = \tfrac{1}{2}\cdot[\hat{\mathbf{x}}_T^*(k) - \tilde{\mathbf{x}}_T^i]'\,\mathbf{Q}_T^i\,[\hat{\mathbf{x}}_T^*(k) - \tilde{\mathbf{x}}_T^i] + \tfrac{1}{2}\cdot\sum_{j=1}^{n}[\mathbf{u}_T^j(k) - \tilde{\mathbf{u}}_T^{ij}]'\,\mathbf{R}_T^{ij}\,[\mathbf{u}_T^j(k) - \tilde{\mathbf{u}}_T^{ij}]. \quad (50)$$

For $\mathbf{P}_T^i(k) := \mathbf{Q}_T^i$ and $\mathbf{p}_T^i(k) := \mathbf{Q}_T^i\tilde{\mathbf{x}}_T^i$, Eq. 50 is equivalent to

$$J_T^i(k) = \tfrac{1}{2}\cdot\hat{\mathbf{x}}_T^{*'}(k)\mathbf{P}_T^i\hat{\mathbf{x}}_T^*(k) - \hat{\mathbf{x}}_T^{*'}(k)\mathbf{p}_T^i(k) + \xi_T(k), \quad (51)$$

where the scalar $\xi_T(k)$ is the sum of all terms that do not depend on $\hat{\mathbf{x}}_T^*(k)$ and $\mathbf{u}_T^i(k)$ and is, thus, without any relevance for our further calculations. From Eq. 5 we know that the optimal state vector for the terminal period can be derived by the use of the state vector optimized for the previous time period, $\hat{\mathbf{x}}_{T-1}^*(k)$, and the optimal control variables, $\hat{\mathbf{u}}_T^{i*}(k)$ $\forall\ i \in \{1,...,n\}$. To derive the latter, in Eq. 51 we replace $\hat{\mathbf{x}}_T^*(k)$ by the right-hand side of Eq. 5, and compute the optimal values of $J_T^{i*}(k)$ $\forall\ i \in \{1,...,n\}$ by minimizing $J_T^i(k)$ with respect to $\mathbf{u}_T^i(k)$, i.e.,

$$\begin{aligned}\frac{\partial J_T^i(k)}{\partial \mathbf{u}_T^i(k)} = \mathbf{B}_T^{i'}(k)\mathbf{P}_T^i(k)[\mathbf{A}_T(k)\hat{\mathbf{x}}_{T-1}^*(k) + \sum_{j=1}^{n}\mathbf{B}_T^j(k)\hat{\mathbf{u}}_T^{j*}(k) + \mathbf{c}_T(k)]\\ -\mathbf{B}_T^{i'}(k)\mathbf{p}_T^i(k) + \mathbf{R}_T^{ii}[\hat{\mathbf{u}}_T^{i*}(k) - \tilde{\mathbf{u}}_T^{ii}] = \mathbf{0}.\end{aligned} \quad (52)$$

Note that, since $J_T^i(k)$ is strictly convex with respect to $\mathbf{u}_T^i(k)$ $\forall\ i \in \{1,...,n\}$, the first-order conditions (Eq. 52) are necessary and sufficient.

Under the assumption that all players act simultaneously, we can derive optimal control variables of the form $\hat{\mathbf{u}}_T^{i*}(k) = \mathbf{G}_T^i(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{g}_T^i(k)$.[14] Plugging these into Eq. 52 we arrive at Eq. 13 and Eq. 14 for $t = T$ respectively, from which we can compute the feedback matrices, $\mathbf{G}_T^i(k)$ and $\mathbf{g}_T^i(k)$. The optimal state can then be determined by $\hat{\mathbf{x}}_T^*(k) = \mathbf{K}_T(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{k}_T(k)$ (cf. Eq. 17 for $t = T$) with $\mathbf{K}_T(k)$ and $\mathbf{k}_T(k)$ given by Eq. 11 and Eq. 12 for $t = T$ respectively.

For the derivation of period-$(T-1)$ parameter matrices of the value function, i.e., the Riccati matrices for time period $T-1$, we set up the cost-to-go function $J_T^{i*}(k) + J_{T-1}^i(k)$ and replace $\hat{\mathbf{x}}_T^*(k)$ and $\hat{\mathbf{u}}_T^{i*}(k)$ by Eq. 17 and Eq. 18 for $t = T-1$ respectively.

$$
\begin{aligned}
J_T^{i*}(k) + J_{T-1}^i(k) &= [\mathbf{K}_T(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{k}_T(k)]' \left[\tfrac{1}{2}\cdot\mathbf{P}_T^i[\mathbf{K}_T(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{k}_T(k)] - \mathbf{p}_T^i(k)\right] \\
&+ \sum_{j=1}^n [\mathbf{G}_T^j(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{g}_T^j(k)]'\mathbf{R}_T^{ij}\left[\tfrac{1}{2}\cdot[\mathbf{G}_T^j(k)\hat{\mathbf{x}}_{T-1}^*(k) + \mathbf{g}_T^j(k)] - \tilde{\mathbf{u}}_T^{ij}\right] \\
&\qquad + \hat{\mathbf{x}}_{T-1}^{*'}(k)\mathbf{Q}_{T-1}^i\left[\tfrac{1}{2}\cdot\hat{\mathbf{x}}_{T-1}^*(k) - \tilde{\mathbf{x}}_{T-1}^i(k)\right] \\
&\quad + \sum_{j=1}^n \mathbf{u}_{T-1}^{j'}(k)\mathbf{R}_{T-1}^{ij}\left[\tfrac{1}{2}\cdot\mathbf{u}_{T-1}^j(k) - \tilde{\mathbf{u}}_{T-1}^{ij}(k)\right] + \psi_{T-1}(k),
\end{aligned}
$$

where the scalar $\psi_{T-1}(k)$ is without any relevance for further calculations since it is the sum of all terms that do not depend on $\hat{\mathbf{x}}_{T-1}^*(k)$ and $\mathbf{u}_{T-1}^i(k)$. Collecting all terms containing $\hat{\mathbf{x}}_{T-1}^*(k)$ we get

$$J_T^{i*}(k) + J_{T-1}^i(k) = \tfrac{1}{2}\cdot\hat{\mathbf{x}}_{T-1}^{*'}(k)\mathbf{P}_{T-1}^i\hat{\mathbf{x}}_{T-1}^*(k) - \hat{\mathbf{x}}_{T-1}^{*'}(k)\mathbf{p}_{T-1}^i(k) + \xi_{T-1}(k), \tag{53}$$

and can identify the Riccati matrices for $T-1$ by comparing coefficients with Eq. 53. The Riccati matrices are then determined by Eq. 9 and Eq. 10 for $t = T-1$ respectively. Then, we minimize the objective function of player i $(i = 1, ..., n)$, i.e.,

$$J_T^{i*}(k) + J_{T-1}^{i*}(k) = \min_{\mathbf{u}_{T-1}^i} \left\{J_T^{i*}(k) + J_{T-1}^i(k)\right\}, \tag{54}$$

analogously to what was done for period T: In Eq. 53 we replace $\hat{\mathbf{x}}_{T-1}^*(k)$ by the linearized system dynamics, $\mathbf{A}_{T-1}(k)\hat{\mathbf{x}}_{T-2}^*(k) + \sum_{i=1}^n \mathbf{B}_{T-1}^i(k)\mathbf{u}_{T-1}^i(k) + \mathbf{c}_{T-1}(k)$, compute the expression's first derivative with respect to $\mathbf{u}_{T-1}^i(k)\ \forall\ i \in \{1, ..., n\}$, and set the derivative equal to zero. A little algebra yields optimal control variables of the form $\hat{\mathbf{u}}_{T-1}^{i*}(k) = \mathbf{G}_{T-1}^i(k)\hat{\mathbf{x}}_{T-2}^*(k) + \mathbf{g}_{T-1}^i(k)$ with $\mathbf{G}_{T-1}^i(k)$ and $\mathbf{g}_{T-1}^i(k)$ being derived by solving $2n$ linear matrix equations consisting of Eq. 13 and Eq. 14 $\forall\ i \in \{1, ..., n\}$ (for $t = T-1$) respectively. The optimal state variable for period $t = T-1$ can, then, be determined by $\hat{\mathbf{x}}_{T-1}^*(k) = \mathbf{K}_{T-1}(k)\hat{\mathbf{x}}_{T-2}^*(k) + \mathbf{k}_{T-1}(k)$ (cf. Eq. 17 for $t = T-1$) with $\mathbf{K}_{T-1}(k)$ and $\mathbf{k}_{T-1}(k)$ given by Eq. 11 and Eq. 12 for $t = T-1$ respectively, with the Riccati matrices being determined by Eq. 9 and Eq. 10 (determined again by comparing coefficients).

The procedure sketched for $t = T$ and $t = T-1$ can be extended to period $t = T-2$ and generalized to any other period $t = \tau$ $(\tau \geq 1)$ by induction. The existence of uniquely determined Riccati matrices for all periods $t \in \{1, ..., T\}$ of the LQDG, i.e., each player i seeking to minimize Eq. 1 subject to Eq. 5, can readily be verified according to, e.g. [1], if the penalty matrices for the states are nonnegative definite (which is what we assumed).

To conclude, the LQDG at iteration step k is solved by starting with the terminal conditions $\mathbf{P}_{iT}(k)$ and $\mathbf{p}_{iT}(k)$, and integrating the Riccati equations (Eqs. 9 and 10) backward in time. Utilizing both Riccati matrices, $\mathbf{P}_t^i(k)$ and $\mathbf{p}_t^i(k)$, and feedback matrices, $\mathbf{G}_t^i(k)$ and $\mathbf{g}_t^i(k)$, computed for all players and for all time periods, i.e., $\forall\ i \in \{1, ..., n\}$

[14] The linear functional form of the optimal control vectors, for $i = 1, ..., n$, results from the quadratic structure of the cost (or payoff) functions (Eq. 1) and the linearity of the system equation (Eq. 5).

and $\forall\ t \in \{1, ..., T\}$, the k^{th} iteration of the feedback Nash equilibrium path for the state variable, $\{\hat{\mathbf{x}}_t^*(k)\}_{t=1}^T$, and the k^{th} iteration of player i's equilibrium path for their own control variable, $\{\hat{\mathbf{u}}_t^{i*}(k)\}_{t=1}^T$, are determined by Eq. 17 and Eq. 18 respectively, both being initiated with $\hat{\mathbf{x}}_0^*(k) = \bar{\mathbf{x}}_0$ (where $\mathbf{K}_t(k)$ and $\mathbf{k}_t(k)$ are defined by Eq. 11 and Eq. 12 respectively).

Given the values of optimal states and controls, the scalar values of the loss functions can be determined.