algorithms

MDPI

Article

Approximately Optimal Control of Nonlinear Dynamic Stochastic Problems with Learning: The OPTCON Algorithm

Dmitri Blueschke *, Viktoria Blueschke-Nikolaeva and Reinhard Neck

Department of Economics, University of Klagenfurt, 9020 Klagenfurt, Austria;
viktoria.blueschke-nikolaeva@aau.at (V.B.-N.); reinhard.neck@aau.at (R.N.)
* Correspondence: dmitri.blueschke@aau.at

Abstract: OPTCON is an algorithm for the optimal control of nonlinear stochastic systems which is particularly applicable to economic models. It delivers approximate numerical solutions to optimum control (dynamic optimization) problems with a quadratic objective function for nonlinear economic models with additive and multiplicative (parameter) uncertainties. The algorithm was first programmed in C# and then in MATLAB. It allows for deterministic and stochastic control, the latter with open loop (OPTCON1), passive learning (open-loop feedback, OPTCON2), and active learning (closed-loop, dual, or adaptive control, OPTCON3) information patterns. The mathematical aspects of the algorithm with open-loop feedback and closed-loop information patterns are presented in more detail in this paper.

Keywords: optimal control; stochastic control; algorithms; open-loop control; open-loop feedback control; closed-loop feedback control; learning

check for updates

Citation: Blueschke, D.; Blueschke-Nikolaeva, V.; Neck, R. Approximately Optimal Control of Nonlinear Dynamic Stochastic Problems with Learning: The OPTCON Algorithm. *Algorithms* **2021**, *14*, 181. https://doi.org/10.3390/a14060181

Academic Editor: Stephanie Allassonniere

Received: 31 March 2021
Accepted: 4 June 2021
Published: 8 June 2021

1. Introduction

Optimal control of stochastic processes is a topic which occurs in many contexts of applied mathematics such as engineering, biology, chemistry, economics, and management science. These techniques are used to steer the behavior of a plant, natural, human or social system over a certain (finite or infinite) time period where the system is driven by differential or difference equations whose time path can be influenced by an external controller. The controller (decision maker, policy maker) aims at achieving the best trajectory of his/her controls according to some criterion (performance measure, objective function) over that time horizon.

Unfortunately, only a few special cases can be solved explicitly, in particular the linear-quadratic-Gaussian problem, where the dynamic system is linear, the objective function is quadratic, and the stochastics are confined to additive normally distributed noise affecting the system dynamics (e.g., [1]). In this case, it is possible to separate the estimation of the state and the optimization of its behavior due to the property of certainty equivalence (the separation theorem). For most practical applications, however, especially those where the parameters of the system are not known precisely, an exact solution to the optimal stochastic control problem is not available and some kind of approximation is required. One of the reasons is the curse of dimensionality, which prevents the solving of dynamic functional equations for multivariable systems [2,3]. This is particularly complicated by the so-called dual effect of controls preventing the separation of estimation and control in adaptive processes; it was discovered by [4] that "the control must have a reciprocal, 'dual' character: to a certain extent they must be of a probing nature, but also controlling to a known degree" ([4], p. 31). The dynamic functional equations for general problems of this kind are known (see, e.g., [5]) but cannot be solved except in very special cases; hence approximations are required (see [6–8]).

In this paper, we present the OPTCON algorithm for calculating numerically (approximately) optimal control solutions to nonlinear dynamic stochastic optimization problems without and with learning. The algorithm was developed in three stages (OPTCON1, OPTCON2, and OPTCON3) over a long period. OPTCON1 determines optimal controls for deterministic problems and stochastic problems based on the open-loop information pattern [9]. OPTCON2 solves problems with passive learning about the unknown parameters of the system model [10], so-called open-loop feedback policies (see [11] for the terminology). Finally, OPTCON3 also includes active learning (dual, adaptive control) according to an approximation procedure initiated by [12–14], and adapted to linear economic systems by [15,16], cf. [17]. The present paper provides mathematical details for OPTCON3, which is the most sophisticated version of the OPTCON algorithm; in [18] we concentrate on computations aspects and applications of OPTCON3.

The OPTCON algorithm is applicable for stochastic control problems with the following properties: The model of the process is multivariable, formulated in discrete time, and described by a system of nonlinear difference equations with known functional form but additive noise and (possibly) unknown parameters. The state is stochastically observable and controllable. No inequality constraints on states or controls are given. Open-loop, open-loop feedback, and closed-loop information structures can be considered. The objective function is quadratic and perfectly known; it is formulated in tracking form but can easily be transformed to the general quadratic form. Quadratic functions can be interpreted as second-order Taylor approximations to more general functional forms. There is only one decision maker; for decentralized problems, see the OPTGAME algorithm [19].

The paper has the following structure. In Section 2 the relevant basic background is provided, namely the class of problems to be solved by the algorithm as well as some basic information about the linear-quadratic framework and nonlinear system solving algorithms. Section 3 presents the OPTCON algorithm stepwise, starting with the basic version (OPTCON1) and then introducing the more advanced versions OPTCON2 and OPTCON3 for handling stochastic components in the optimal control process. In Section 4 some details on computational time and accuracy of the OPTCON algorithm are presented. Section 5 concludes.

2. Theoretical Background

2.1. Optimal Control Problem

We consider optimal control problems with a quadratic objective function and a nonlinear multivariate discrete-time dynamic system under additive and parameter uncertainties. The basis for an optimal control problem is a deterministic dynamic system to be controlled (plant, firm, economy, etc.) in discrete time in the form:

$$x_t = f(x_{t-1}, x_t, u_t, z_t), \quad t = 1, ..., T, \tag{1}$$

where:

- $x_t \in R^n$ is a vector of state variables that describes the state of the system at any point in time t,
- $u_t \in R^m$ is a vector of control variables; we assume that the decision maker determines the values of the control variables exactly (without randomization) according to the approximately optimal solution of the problem,
- $z_t \in R^l$ denotes a vector of non-controlled deterministic exogenous variables, whose values are known to the decision maker at time t,
- T denotes the terminal time period of the finite planning horizon.

The function f is assumed to be twice continuously differentiable with respect to all of its arguments.

We assume that there is some true law of motion given by Equation (1) in the background which is at least partially unknown to the policy maker while the function f is known to him/her. The policy maker faces two sources of uncertainty, additive and parameter uncertainties where:

- $\theta \in R^p$ denotes a vector of stochastic variables of system parameters (parameter uncertainty),
- $\overline{\theta} \in R^p$ is a vector of true parameters whose values are assumed to be constant but unknown to the policy maker,
- $\varepsilon_t \in R^n$ is a vector of additive stochastic disturbances (system error).

θ_t and ε_t are assumed to be independent Gaussian random vectors with expectations $\hat{\theta}$ and O_n respectively and covariance matrices $\Sigma^{\theta\theta}$ and $\Sigma^{\varepsilon\varepsilon}$ respectively.

Including uncertainty, the system (1) can be written as:

$$x_t = f(x_{t-1}, x_t, u_t, \theta_t, z_t) + \varepsilon_t, \quad t = 1, ..., T. \tag{2}$$

The optimal control problem to be solved approximately assumes that there is a modeler who wishes to control the system (2). It means that the controller wishes to bring the state variables, using the control variables, as close as possible to some pre-defined desired time path, according to a given optimality criterion. The OPTCON algorithm allows for the optimal control of the system (2) using a quadratic objective function. To this end the modeler needs to define the following variables:

- $\tilde{x}_t \in R^n$ are given target ('ideal') values (for $t = 1, ..., T$) of the state variables,
- $\tilde{u}_t \in R^m$ are given target ('ideal') values (for $t = 1, ..., T$) of the control variables,
- W_t is an $((n + m) \times (n + m))$ symmetric matrix defining the relative weights of the state and control variables in the objective function. In a frequent special case, W_t includes a discount factor α with $W_t = \alpha^{t-1}W$.

The resulting intertemporal objective function is given in quadratic tracking form:

$$J = E\left[\sum_{t=1}^{T} L_t(x_t, u_t)\right], \tag{3}$$

with

$$L_t(x_t, u_t) = \frac{1}{2}\begin{pmatrix} x_t - \tilde{x}_t \\ u_t - \tilde{u}_t \end{pmatrix}' W_t \begin{pmatrix} x_t - \tilde{x}_t \\ u_t - \tilde{u}_t \end{pmatrix}. \tag{4}$$

The OPTCON algorithm allows us to find approximate optimal control solutions which minimize the functions (3) and (4) and satisfy the system dynamics (2).

2.2. Linear-Quadratic Optimal Control (LQ) Framework

The novel feature of the OPTCON algorithm is the combination of a nonlinear dynamic system and the above-mentioned stochastics when solving optimal control problems. Before we can discuss this technique in detail, a few words should be said about the underlying linear-quadratic optimal control framework. The dynamic system is given in linear form as

$$x_t = Ax_{t-1} + Bu_t + c_t. \tag{5}$$

The corresponding optimal control objective function can be written in the same form as in (3) and (4). Using well-known LQ optimization techniques, the rules for defining control variables can be found iteratively backwards in time using dynamic Riccati equations. We start from the familiar LQ framework (e.g., [20]), which was adapted to economic models by [15,21,22]. The same standard technique is applied in the OPTCON algorithm whenever an optimal control for a linearized system should be obtained.

Furthermore, for computational reasons it is useful to transfer the objective function from the quadratic tracking form as used in Equation (4) into the general quadratic form:

$$L_t(x_t, u_t) = \frac{1}{2}\begin{pmatrix} x_t \\ u_t \end{pmatrix}' W_t \begin{pmatrix} x_t \\ u_t \end{pmatrix} + \begin{pmatrix} x_t \\ u_t \end{pmatrix}' \begin{pmatrix} w_t^x \\ w_t^u \end{pmatrix} + w_t^c, \tag{6}$$

where

$$\begin{pmatrix} w_t^x \\ w_t^u \end{pmatrix} = -W_t \begin{pmatrix} \tilde{x}_t \\ \tilde{u}_t \end{pmatrix} \; and \; w_t^c = \frac{1}{2} \begin{pmatrix} \tilde{x}_t \\ \tilde{u}_t \end{pmatrix}' W_t \begin{pmatrix} \tilde{x}_t \\ \tilde{u}_t \end{pmatrix}.$$

The equivalence between the quadratic tracking form and the general quadratic form is shown, for instance, in [23].

2.3. Solving Nonlinear Dynamic Systems

The OPTCON algorithm allows us to find approximately optimal control solutions for nonlinear stochastic systems. In the solution process, the nonlinear dynamic system (2) should be solved in a large number of iterations. For this reason the choice of an appropriate solver is important to guarantee obtaining reliable solutions taking the computational costs into account. In the OPTCON algorithm, different solution algorithms are included attaching different importance to the trade-off between reliability and computational speed. At the moment the following methods can be used: Levenberg–Marquardt [24], trust region [25], Newton–Raphson [26], or Gauss–Seidel [27].

3. The OPTCON Algorithm

3.1. The OPTCON1 Algorithm

The first version of the OPTCON algorithm, the OPTCON1 algorithm, was developed by [9]. It allows us to calculate an open-loop (OL) solution to a nonlinear stochastic dynamic optimal control problem with a quadratic objective function under additive and parameter uncertainties. Open-loop controls either do not take account of the effect of uncertainties in the system or assume the stochastics (expectation and covariance matrices of additive and multiplicative disturbances) to be given for all time periods at the beginning of the planning horizon. The basic idea behind this algorithm is that it extends linear-quadratic stochastic optimal control techniques (see, e.g., [1]) to nonlinear problems using an iterative method of linear approximations.

The OPTCON algorithm solves the nonlinear optimal control problem iteratively by a sequence of linear approximations where the next approximation is derived using the optimal control solution of the previous one. The (optimal) solutions of the intermediate linear approximations are iterated from one time path to another until the algorithm converges. The criterion for convergence is that the difference in the values of the state and control variables of current and previous iterations is smaller than a pre-specified number. or the maximal number of iterations is reached. In each iteration the same procedure is conducted, namely linearizing and optimizing the linearized system. The system is linearized around the previous iteration's result and the problem is solved for the resulting time-varying linearized system. The solution of the linearized model is obtained via Bellman's principle of optimality and is used as the tentative path for the next iteration, starting off the procedure all over again. If the OPTCON algorithm converges, the solution of the last iteration is taken as the approximately optimal solution to the nonlinear problem and the value of the objective function is calculated. Figure 1 shows the structure of the OPTCON1 algorithm.

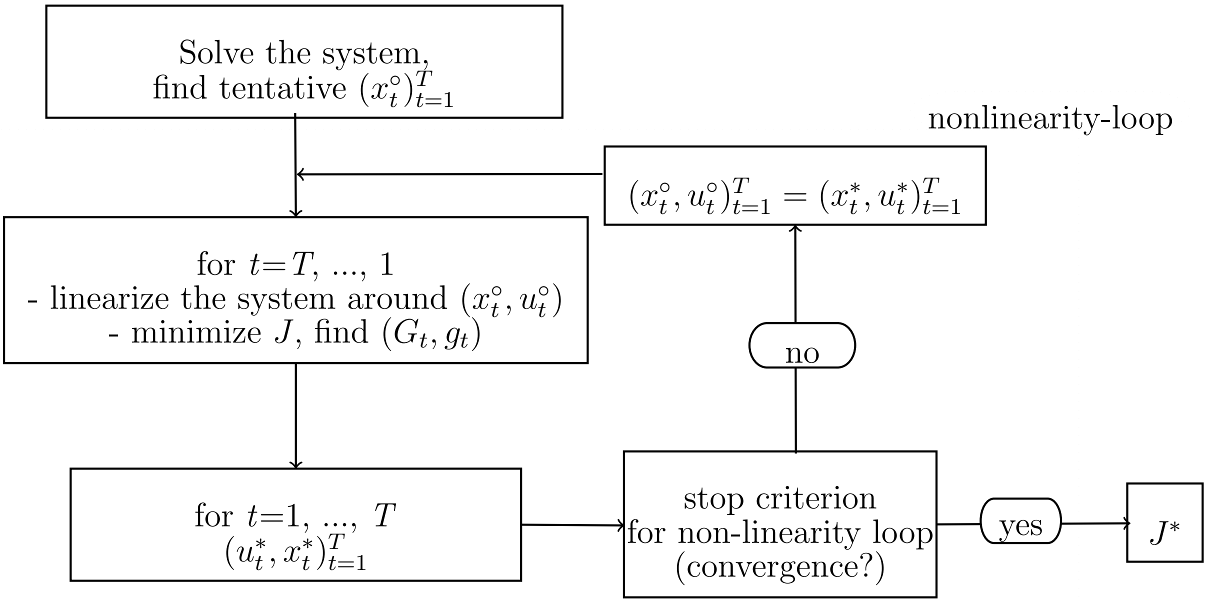
Solve the system,
find tentative $(x_t^\circ)_{t=1}^T$
nonlinearity-loop
$(x_t^\circ, u_t^\circ)_{t=1}^T = (x_t^*, u_t^*)_{t=1}^T$
for $t=T, ..., 1$
- linearize the system around (x_t°, u_t°)
- minimize J, find (G_t, g_t)
no
for $t=1, ..., T$
$(u_t^*, x_t^*)_{t=1}^T$
stop criterion
for non-linearity loop
(convergence?)
yes
J^*

Figure 1. Flow chart of OPTCON1.

3.2. The OPTCON2 Algorithm

The second version of the algorithm (called OPTCON2, see [10,23] for more details) extends the first version with a passive learning strategy (open-loop feedback (OLF)). OLF uses the idea of re-estimating the model at the end of each time period $t = 1, ..., T$. For that purpose, the model builder (policy maker) observes the current situation and uses the new information, namely the current values of the state variables, to improve his/her knowledge of the system.

Recall that the stochastic system includes two kinds of uncertainties, namely additive (random system errors) and multiplicative ('structural' errors in parameters). We assume that there is some true law of motion in the background which is unknown to the policy maker. The passive learning strategy deals with the 'true' parameters $\overline{\theta}$ which generate the model. However, the policy maker does not know these true parameters $\overline{\theta}$ and works with the 'wrong' estimated parameters $\hat{\theta}$. The idea of the OLF strategy is to observe the outcome of the model in each time period and use this information to bring the estimated parameters $\hat{\theta}$ closer to the true values $\overline{\theta}$.

The OLF strategy of the OPTCON2 algorithm consists of the following (rough) steps running in a forward loop.

In each time period S ($S = 1, \ldots, T$) do the following:

- Find an (approximately) optimal open-loop solution for the remaining subproblem (for the time periods from S to T) using the OPTCON1 algorithm.
- Fix the predicted solution for time period S, namely x_S^* and u_S^*.
- Observe the realized state variables x_S^{a*} which result from the chosen control variables (u_S^*), the true law of motion with parameter vector $\overline{\theta}$, and the realization of the random disturbances ε_S. The main difference between x_S^{a*} and x_S^* is that the former is driven by the true dynamic process with parameter vector $\overline{\theta}$ and the latter by the estimated dynamic system with $\hat{\theta}$.
- Update the parameter estimate $\hat{\theta}$ (via the Kalman filter and using the difference between x_S^{a*} and x_S^*) and use it in the next iteration step $S + 1$.

The last step is carried out using the idea of the Kalman filter (see, e.g., [28] or [29]). The updating procedure according to the Kalman filter consists of two distinct phases, prediction and correction. First (the prediction phase), the predicted values of the state variables $\hat{x}_{S/S-1}$, the vector of parameters $\hat{\theta}_{S/S-1}$, and the covariances of the parameters $\Sigma_{S/S-1}^{\theta\theta}$

are calculated using the estimates from the previous time period. Next (the correction phase), these 'a priori' values of the state variables, the vector of the parameters, and the values of the covariances are corrected using the current observations of the state variables.

The phase of calculating the predicted values of $\hat{x}_{S/S-1}$ and $\hat{\theta}_{S/S-1}$ is integrated in the previous steps of the OPTCON2 algorithm. At the end of time period S the predicted values of $\hat{x}_{S/S-1} = x^*_S$ are calculated and $\hat{\theta}_{S/S-1} = \hat{\theta}_{S-1}$ is known.

The correction phase is reduced to the calculation of the parameter estimates $\hat{\theta}_{S/S}$ and the parameter covariances $\Sigma^{\theta\theta}_{S/S}$ because the 'corrected' values of the state variables $x^{a*}_S = f(x^{a*}_{S-1}, x^{a*}_S, u^*_S, \overline{\theta}, z_S) + \varepsilon_S$ are calculated in the previous step of the algorithm or simply observed. As explained before, the updating procedure is based on the difference between $x^*_S = f(x^{a*}_{S-1}, x^*_S, u^*_S, \hat{\theta}, z_S)$ and x^{a*}_S at the end of each time period.

The OPTCON2 algorithm allows us to find an approximate solution of the optimal control problem using a passive learning strategy. The policy maker uses real observations in each time period to update his/her knowledge about the stochastic dynamic system. In many cases it allows him/her to obtain more reliable results (see [30] for a performance study on optimal control with a passive learning strategy).

3.3. The OPTCON3 Algorithm

In this section we present the detailed description of the OPTCON3 algorithm. As mentioned above, the latest version of the OPTCON algorithm includes an active learning strategy. In the literature this strategy is also called closed-loop, adaptive, or dual control. The passive learning method in the OPTCON2 algorithm uses current observations to update the information about the dynamic system. The active learning strategy in OPTCON3 actively uses system perturbations to improve the optimal control performance by reducing the uncertainty in the future.

The procedure for finding the closed-loop solution in the OPTCON3 algorithm is based on the linear active learning control method in [15]. Similar to the iterative structure in the OPTCON2 algorithm, the optimization runs in a forward loop ($S = 1, \ldots, T$). However, in each time period, more information about future measurements is used. To this end the objective function J (as defined in Equations (3) and (4)) is extended for stochastic components. We define in each time period J_d as the approximate total cost-to-go with $T - S$ periods remaining. The approximate cost-to-go is broken down into three terms: $J_d = J_D + J_C + J_P$. J_D is called deterministic term and incorporates only non-stochastic components. J_C is the cautionary factor that includes the stochastic elements in the current period. J_P is called probing term and captures the effect of dual learning on the uncertainty in the future periods. J_C and J_P constitute a separate quadratic minimization problem constrained by the nonlinear system. The original system needs to be expanded to the perturbation form δx_t. The optimization problem has to be solved for the perturbed system, where ΔJ^*_t is the perturbed objective function. Using Taylor expansion of nonlinear system and Bellman optimization technique, we obtain the solution ΔJ^*_t as a quadratic function of δx_{t-1}. The original J^*_t can be derived from ΔJ^*_t. To take uncertain parameters into account, all the terms and formulas need to be adjusted to the augmented system $\begin{bmatrix} x_t \\ \ldots\ldots \\ \theta \end{bmatrix}$.

Next, a schematic structure of the OPTCON3 algorithm is presented. This structure is visualized in Figure 2.

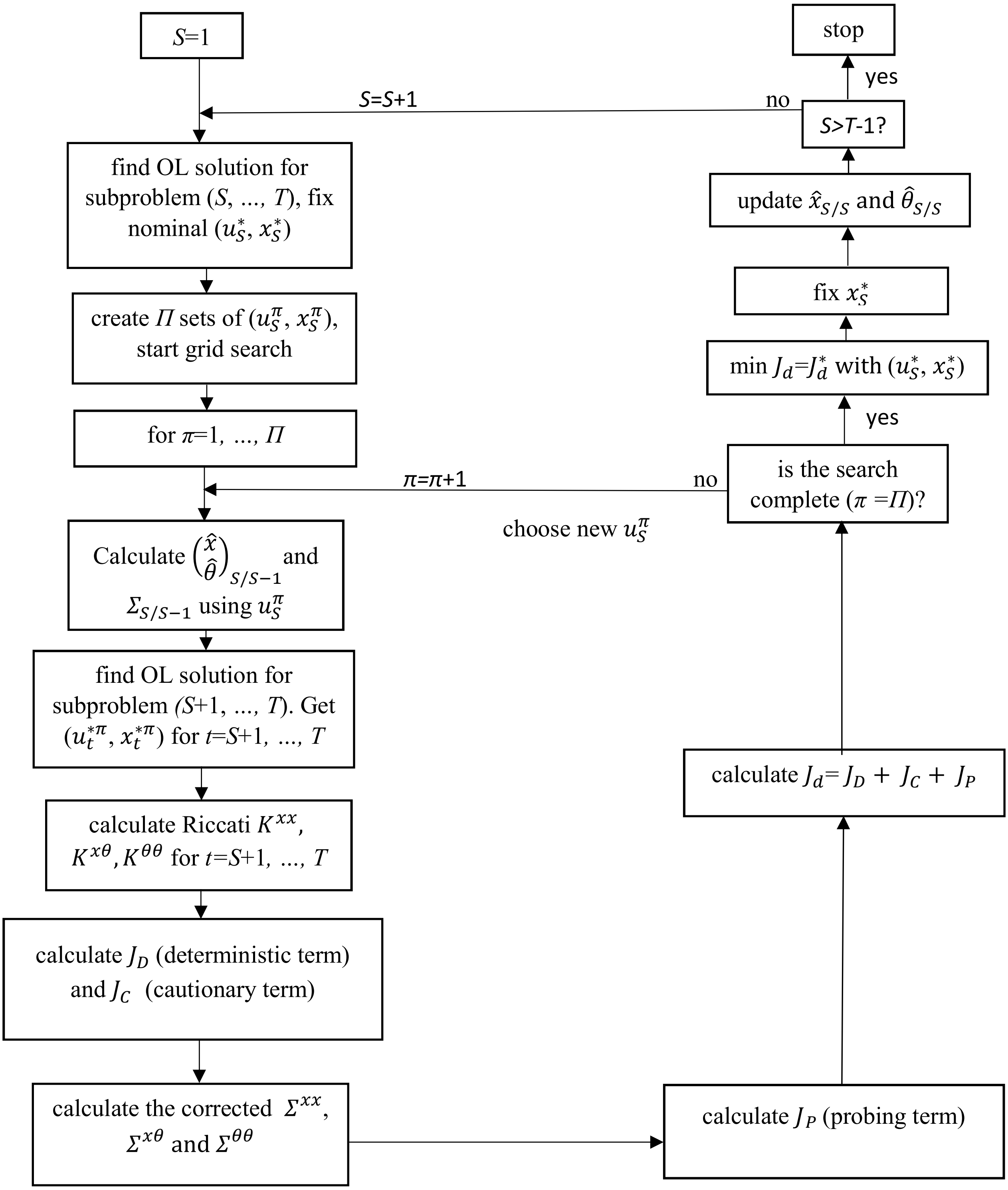
S=1
stop
yes
S=S+1
no
S>T-1?
find OL solution for subproblem (S, ..., T), fix nominal (u_S^*, x_S^*)
update $\hat{x}_{S/S}$ and $\hat{\theta}_{S/S}$
fix x_S^*
create Π sets of (u_S^π, x_S^π), start grid search
min J_d=J_d^* with (u_S^*, x_S^*)
for π=1, ..., Π
yes
π=π+1
no
is the search complete (π =Π)?
choose new u_S^π
Calculate $\begin{pmatrix}\hat{x}\\ \hat{\theta}\end{pmatrix}_{S/S-1}$ and $\Sigma_{S/S-1}$ using u_S^π
find OL solution for subproblem (S+1, ..., T). Get ($u_t^{*\pi}$, $x_t^{*\pi}$) for t=S+1, ..., T
calculate J_d=J_D + J_C + J_P
calculate Riccati K^{xx}, $K^{x\theta}$,$K^{\theta\theta}$ for t=S+1, ..., T
calculate J_D (deterministic term) and J_C (cautionary term)
calculate the corrected Σ^{xx}, $\Sigma^{x\theta}$ and $\Sigma^{\theta\theta}$
calculate J_P (probing term)

Figure 2. Flow chart of OPTCON3.

The optimization is carried out in a forward loop from 1 to T conducting following procedure in each time period S ($S = 1, \ldots, T$). The subproblem from S to T is solved via the open-loop (OL) strategy (see Figure 1 in Section 3.1). The OL solution (x_S^*, u_S^*) for the time period S is fixed. After that the core part of the dual control strategy starts, where the system is actively perturbed to learn about it in order to minimize the uncertainty and the objective function values in the future. In the OPTCON3 algorithm a grid search method is used (instead of grid search many other methods can be applied, such as gradient optimization or a more advanced heuristic approach). In the so-called "π-loop" we create a grid of possible solutions around the existing path (x_S^*, u_S^*). In each iteration ($\pi = 1, ..., \Pi$), i.e., in each grid point, the optimal control solution is obtained and evaluated using the objective function. Inside the π-loop (for each π), the following steps are carried out.

An optimal open-loop solution for the subproblem from $S+1$ to T is determined. Then the OL solution $(x_{S+1}^{*\pi}, u_{S+1}^{*\pi})$ for the time period $S+1$ is fixed. Next, after some auxiliary calculations (including, among other, Riccati matrices), the deterministic J_D, cautionary J_C, and probing J_P terms of the cost-to-go are determined in a forward loop from $S+1$ to T. Once the iteration of the π-loop has been completed, the total approximate objective function $J_d = J_D + J_C + J_P$ can be obtained. When the search is completed, i.e., the approximately optimal path with $min J_d$ has been found, the obtained optimal control in time period $S+1$ is carried out by the policy maker. As a result of these controls and the true dynamic systems the actual values of the state variables can be obtained. This new information is used by the policy maker to update and to adjust the parameter estimate $\hat{\theta}$, whereby the Kalman filter is used. Using the updated stochastic parameters, the same active learning procedure is applied for the remaining time periods from $S+2$ to T.

The OPTCON3 algorithm basically uses the technique presented in [13,15] but is augmented by approximating, in each step, the nonlinear system in a series of linear systems (iterative procedure as described in Section 3.1).

The following steps (I–IV) of the OPTCON3 algorithm describe how to obtain an approximately optimal dual control solution to a stochastic problem.

The algorithm needs the following input values (A legitimate question would be how to obtain the required (tentative) inputs. It obviously depends on the research question. In the case of some numerical experiments, one can resort to Monte Carlo simulations. In the case of a real-world application, one can estimate system parameters by an appropriate method or simply guess some values):

f	system function
$x_0 = \overset{\circ}{x}_0$	initial values of state variables
$(\overset{\circ}{u}_t)_{t=1}^T$	tentative path of control variables
$\hat{\theta}_0 = \hat{\theta}$	expected values of system parameters
$\Sigma_0^{\theta\theta} = \Sigma^{\theta\theta}$	covariance matrix of system parameters
$\Sigma^{\varepsilon\varepsilon}$	covariance matrix of system noise
$(\varepsilon_t)_{t=1}^T$	system noises
$(z_t)_{t=1}^T$	path of exogenous variables
$(\tilde{x}_t)_{t=1}^T$	target path for state variables
$(\tilde{u}_t)_{t=1}^T$	target path for control variables
W^{xx}, W^{ux}, W^{uu}	weighting matrices of objective function
w_t^u, w_t^x	weighting vectors of objective function
α	discount factor of objective function.

As a result of the algorithm, an approximately optimal solution $(x_t^{a*})_{t=1}^T$, $(u_t^*)_{t=1}^T$ and the corresponding value of the objective function J^* are obtained.

Step I: Do the following search steps [1], [2] and [3] for each $S = 1, ..., T$:

Step I-[1]: Find the OL solution for the subproblem $(S, ..., T)$: apply the procedure already used in OPTCON1. Fix (x_S^*, u_S^*).

Step I-[2]: Run a grid search of size Π around (x_S^*, u_S^*), i.e., perform steps (A)–(G) for each $\pi = 1, ..., \Pi$:

Step I-2A: Find the OL solution $(x_t^{*\pi}, u_t^{*\pi})_{S+1}^T$ for the subproblem $(S+1, ..., T)$.

• The nonlinearity loop is run until the stop criterion is fulfilled (the stop criterion is fulfilled when the difference between the values of the current and the previous iteration is smaller than a pre-specified number or the maximum number of iterations is achieved). As a result the approximately optimal solution $(x_t^{*\pi}, u_t^{*\pi})_{S+1}^T$ has been found. Then go to the next step I-2B.

Notes:

- After several runs of the nonlinearity loop, only the solution $(x^{*\pi}_{S+1}, u^{*\pi}_{S+1})$ for the time period $S+1$ will be taken as the optimal (nominal) solution. The calculations of the pairs $(x^{*\pi}_{t'}, u^{*\pi}_{t'})$ for other periods ($t' > S+1$) have to be done again, taking into account the re-estimated parameters for all periods.
- After linearization the parameter matrices for the linearized system are obtained: $A_t = (I - F^x_{x_t})^{-1} F^x_{x_{t-1}}$ and $B_t = (I - F^x_{x_t})^{-1} F^x_{u_t}$, where $F^x_{x_t}$, $F^x_{x_{t-1}}$, and $F^x_{u_t}$ are the derivatives of the system function f with respect to x_t, x_{t-1}, and u_t respectively.

Step I-2B: Do a backward recursion to obtain auxiliary matrices for calculating components of the objective function.

Initialize the following auxiliary matrices for backward recursion:

$$H^{xx}_{T+1} = O_{n\times n}, h^x_{T+1} = O_n,$$
$$H^{x\theta}_{T+1} = O_{n\times p}, H^{\theta\theta}_{T+1} = O_{p\times p}.$$

Calculate the Riccati matrices K^{xx}_t, $K^{\theta x}_t$, and $K^{\theta\theta}_t$ and the auxiliary matrices Λ^{xx}_t, Λ^{xu}_t, Λ^{uu}_t, λ^x_t, λ^u_t, H^{xx}_t, $H^{\theta x}_t$, $H^{\theta\theta}_t$, and h^x_t backwards in time from T to $S+1$.

$$\begin{aligned}&K^{xx}_t = W^{xx}_t + H^{xx}_{t+1}, K^{\theta x}_t = H^{\theta x}_{t+1}, K^{\theta\theta}_t = H^{\theta\theta}_{t+1},\\ &k^x_t = h^x_{t+1} + x'W^{xx}_t + W^{xu}_t u + w^x_t.\end{aligned} \tag{7}$$

$$\begin{aligned}\Lambda^{xx}_t &= (A_t)' K^{xx}_t A_t,\\ \Lambda^{ux}_t &= (B_t)' K^{xx}_t A_t + W^{ux}_t A_t,\\ \Lambda^{xu}_t &= (\Lambda^{ux}_t)',\\ \Lambda^{uu}_t &= (B_t)' K^{xx}_t B_t + 2(B_t)' W^{xu}_t + W^{uu}_t.\end{aligned} \tag{8}$$

$$\begin{aligned}\lambda^x_t &= k^x_t A_t,\\ \lambda^u_t &= k^x_t B_t + x'W^{xu}_t + u'W^{uu}_t + (w^u_t)'.\end{aligned} \tag{9}$$

$$\begin{aligned}H^{xx}_t &= \Lambda^{xx}_t - \Lambda^{xu}_t (\Lambda^{uu}_t)^{-1} \Lambda^{ux}_t,\\ H^{\theta x}_t &= [D'K^{xx}_t + K^{\theta x}_t] A_t - [[D'K^{xx}_t + K^{\theta x}_t] B_t + DW^{xu}] (\Lambda^{uu}_t)^{-1} \Lambda^{ux}_t,\\ H^{\theta\theta}_t &= D'(K^{xx}_t D + K^{x\theta}_t) + K^{\theta x}_t D + K^{\theta\theta}_t - [[D'K^{xx}_t + K^{\theta x}_t] B_t + DW^{xu}]\\ &\quad \times (\Lambda^{uu}_t)^{-1} [B_t [K^{xx}_t D + K^{x\theta}_t] + W^{ux} D],\\ h^x_t &= \lambda^x_t - \Lambda^{xu}_t (\Lambda^{uu}_t)^{-1} \lambda^u_t,\end{aligned} \tag{10}$$

where $D = (I - F^x_{x_t})^{-1} F^x_\theta$.
F^x_θ is the derivative of the system function with respect to θ.

Step I-2C: Calculate the deterministic component of the approximate objective function $J_{D,S}$ and the cautionary component $J_{C,S}$:

$$\begin{aligned}J_{D,S} &= \tfrac{1}{2}[x^{*\pi}_S - \tilde{x}_S]' W^{xx} [x^{*\pi}_S - \tilde{x}_S] + [x^{*\pi}_S - \tilde{x}_S]' W^{xu} [u^{*\pi}_S - \tilde{u}_S]\\ &\quad + \tfrac{1}{2}[u^{*\pi}_S - \tilde{u}_S]' W^{uu} [u^{*\pi}_S - \tilde{u}_S]\end{aligned} \tag{11}$$

and

$$J_{C,S} = \tfrac{1}{2} tr(H^{xx}_{S+1} \Sigma^{xx}_{S+1/S}) + tr(H^{\theta x}_{S+1} \Sigma^{x\theta}_{S+1/S}) + \tfrac{1}{2} tr(H^{\theta\theta}_{S+1} \Sigma^{\theta\theta}_{S+1/S}).$$

Step I-2D: Repeat steps [a]–[c] for each $j = S+1, ..., T$:

[a]: Calculate the components $J_{D,j}$ and $J_{C,j}$:

$$\begin{aligned} J_{D,j} &= \tfrac{1}{2}[x_j^{*\pi} - \tilde{x}_j]' W^{xx} [x_j^{*\pi} - \tilde{x}_j] + [x_j^{*\pi} - \tilde{x}_j]' W^{xu} [u_j^{*\pi} - \tilde{u}_j] \\ &\quad + \tfrac{1}{2}[u_j^{*\pi} - \tilde{u}_j]' W^{uu} [u_j^{*\pi} - \tilde{u}_j]), \\ J_{C,j} &= \tfrac{1}{2} tr(K_j^{xx} \Sigma^{\xi\xi}). \end{aligned} \tag{12}$$

[b]: Calculate the matrix $\Sigma^{\theta\theta}_{j/j}$:

$$\Sigma^{\theta\theta}_{j/j} = \Sigma^{\theta\theta}_{j/j-1} - \Sigma^{\theta x}_{j/j-1} (\Sigma^{xx}_{j/j-1})^{-1} \Sigma^{x\theta}_{j/j-1} \tag{13}$$

with

$$\begin{aligned} &\Sigma^{xx}_{j/j-1} = F^x_\theta \Sigma^{\theta\theta}_{j-1/j-1} (F^x_\theta)' + \Sigma^{\varepsilon\varepsilon}_j \\ &\text{and} \\ &\Sigma^{x\theta}_{j/j-1} = (\Sigma^{\theta x}_{j/j-1})' = F^x_\theta \Sigma^{\theta\theta}_{j-1/j-1}, \quad \Sigma^{\theta\theta}_{j/j-1} = \Sigma^{\theta\theta}_{j-1/j-1}. \end{aligned} \tag{14}$$

[c]: Compute the probing component J_{P_j}:

$$\begin{aligned} J_{P_j} &= \tfrac{1}{2} tr[\Lambda^{xu} (\Lambda^{uu})^{-1} \Lambda^{ux} \Sigma^{xx}_{j/j}] \\ &\quad + tr[\Lambda^{xu} (\Lambda^{uu})^{-1} (B'(K^{xx}D + K^{x\theta}) + W^{ux}D) \Sigma^{\theta x}_{j/j}] \\ &\quad + \tfrac{1}{2} tr[((K^{\theta x} + D'K^{xx})B + D'W^{xu})(\Lambda^{uu})^{-1} \\ &\quad \times (B'(K^{xx}D + K^{x\theta}) + W^{ux}D) \Sigma^{\theta\theta}_{j/j}]. \end{aligned} \tag{15}$$

Step I-2E: Calculate the sum of the deterministic, cautionary, and probing terms over the periods $S, ..., T$:

$$J_d = (J_{D,S} + \sum_{j=S+1}^{T} J_{D,j}) + (J_{C,S} + \sum_{j=S+1}^{T} J_{C,j}) + \sum_{j=S+1}^{T} J_{P_j}.$$

Step I-2F: Take a new control $\overset{\circ}{u}_S = \overset{\circ}{u}_S^{\pi+1}$ (a new point in the grid search) and go to step I-2A.

Step I-[3]: Choose an optimal u_S^* with $minJ = J_d^*(u_S^*)$. End of grid search. Fix the corresponding x_S^*.

Step II: Calculate the following (a) and (b) for only one time period S:

(a)

$$\begin{aligned} &\Sigma^{xx}_{S/S-1} = F^x_\theta \Sigma^{\theta\theta}_{S-1/S-1} (F^x_\theta)' + \Sigma^{\varepsilon\varepsilon}_S \\ &\text{and} \\ &\Sigma^{x\theta}_{S/S-1} = (\Sigma^{\theta x}_{S/S-1})' = F^x_\theta \Sigma^{\theta\theta}_{S-1/S-1}, \quad \Sigma^{\theta\theta}_{S/S-1} = \Sigma^{\theta\theta}_{S-S/S-1}. \end{aligned} \tag{16}$$

(b)

$$x_S^{a*} = f(x^{a*}_{S-1}, x_S^{a*}, u_S^*, \overline{\theta}, z) + \varepsilon_S.$$

Step III: Update the parameter estimates $\hat{\theta}$ and $\Sigma^{\theta\theta}_{S/S}$:

$$\hat{\theta}_{S/S} = \hat{\theta}_{S/S-1} + \Sigma^{\theta x}_{S/S-1} (\Sigma^{xx}_{S/S-1})^{-1} [x_S^{a*} - x_S^*] \quad \text{and} \quad \hat{x}_{S/S} = x_S^{a*}. \tag{17}$$

$$\Sigma^{\theta\theta}_{S/S} = \Sigma^{\theta\theta}_{S/S-1} - \Sigma^{\theta x}_{S/S-1} (\Sigma^{xx}_{S/S-1})^{-1} \Sigma^{x\theta}_{S/S-1}. \tag{18}$$

Step IV: Set $\hat{\theta} = \hat{\theta}_{S/S}$ and $\Sigma^{\theta\theta} = \Sigma^{\theta\theta}_{S/S}$, go to Step I and run the procedure for the time period $S + 1$.

The OPTCON3 algorithm is finished when $S = T$ and the approximately optimal dual control and state variables have been found for all periods.

3.4. Computational Details for the AL Procedure

This subsection includes the technical computations for the AL procedure of the OPTCON3 algorithm.

The basic tasks are applying Bellman's principle of optimality, linearization of the nonlinear system in perturbation form, minimizing the objective function, and inserting the obtained feedback rule for controls back into the objective function. In this way the (temporary) components of the optimal objective function can be derived. After that, due to the stochastic nature of the parameters (θ), all the components have to be adjusted to the extended system $\begin{bmatrix} x_t \\ \dots\dots \\ \theta \end{bmatrix}$.

Start with Bellman's principle of optimality:

$$J^*_{T-t} = \min_{u_t} E\{L_t(x_t, u_t) + J^*_{T-(t+1)}(x_t)\}. \tag{19}$$

The last term of (19) is

$$J^*_{T-(t+1)}(x_t) = \min_{u_{t+1}} E\left\{ \dots \min_{u_{T-1}} E\{ \sum_{j=t+1}^{T} L_j(x_j, u_j)/P^{T-1}\}\dots/P^{t+1}\} \right\},$$

where $P^t = (\hat{x}_{t/t}, \Sigma_{t/t})$.

For $\sum_{j=t+1}^{T} L_j(x_j, u_j)$ use the (Taylor) series expansion:

$$\sum_{j=t+1}^{T} L_j(x_j, u_j) = \sum_{j=t+1}^{T} L_j(x_{oj}, u_{oj}) + \sum_{j=t+1}^{T} (L_{jx}\delta x_j + \frac{1}{2}\delta x_j' L_{jxx}\delta x_j$$
$$+\delta x_j' L_{jxu}\delta u_j + L_{ju}\delta u_j + \frac{1}{2}\delta u_j' L_{juu}\delta u_j),$$

where L_{jx}, L_{ju} are the gradients and L_{jxx}, L_{jxu}, and L_{juu} are the Hessian matrices; $\delta x_j = x_j - x_{oj}$, $\delta u_j = u_j - u_{oj}$, and (x_{oj}, u_{oj}) is a nominal path.

Thus,

$$J^*_{T-(t+1)}(x_t) = J^*_{o,T-(t+1)}(x_t) + \Delta J^*_{T-(t+1)}(x_t) \tag{20}$$

and

$$\begin{aligned}\Delta J^*_{T-(t+1)}(x_t) = \min_{u_{t+1}} E\{\dots \min_{u_{T-1}} E\{\textstyle\sum_{j=t+1}^{T}(L_{jx}\delta x_j + \frac{1}{2}\delta x_j' L_{jxx}\delta x_j \\ +\delta x_j' L_{jxu}\delta u_j + L_{ju}\delta u_j + \tfrac{1}{2}\delta u_j' L_{juu}\delta u_j/P^{T-1}\}\dots/P^{t+1})\}.\end{aligned} \tag{21}$$

Next, apply the Taylor expansion to the nonlinear system, i.e., linearize the system function f in (2) around the reference values:

$$x_t \approx f_t(\hat{x}_{t-1/t-1}, u_t^*, x_t) + F_{x_{t-1}}(x_{t-1} - \hat{x}_{t-1/t-1}) + F_u(u_t - u_t^*) + F_{x_t}(x_t - \hat{x}_{t/t}) + \varepsilon_t.$$

This can be rewritten in perturbation form:

$$\delta x_t = F_{x_{t-1}}\delta x_{t-1} + F_u\delta u_t + F_{x_t}\delta x_t + \varepsilon_t$$

After transformation

$$\delta x_t = (I - F_{x_t})^{-1} F_{x_{t-1}}\delta x_{t-1} + (I - F_{x_t})^{-1} F_{u_t}\delta u_t + (I - F_{x_t})^{-1}\varepsilon_t. \tag{22}$$

Thus, (21) and (22) constitute the problem in perturbation form.

Next, we have to prove that the term (21) can be expressed in the quadratic form as a function of δx_t only.
Assume

$$\Delta J^*_{T-(t+1)}(x_t) = g_{t+1} + E\{h^x_{t+1}\delta x_t + \frac{1}{2}\delta x_t' H_{t+1}\delta x_t / P^{t+1}\} \tag{23}$$

and apply the method of induction. By doing so the parameters g, h^x and H are determined.

The next step is to prove that rule (23) is also true for $\Delta J^*_{T-j}(x_t)$ $\forall j$.
Start with

$$J^*_{T-j} = \min_{u_j} E\{L_j(x_j, u_j) + J^*_{T-(j+1)}(x_j)\}.$$

By second order expansion we get:

$$\begin{aligned} J^*_{T-j} = &\min_{u_j} E\{L_j(x_{oj}, u_{oj}) + L_x\delta x_j + \frac{1}{2}\delta x_j' L_{xx}\delta x_j + \delta x_j' L_{xu}\delta u_j + L_u\delta u_j \\ &+ \frac{1}{2}\delta u_j' L_{uu}\delta u_j + (J^*_{o,T-(j+1)} + \Delta J^*_{T-(j+1)}) \end{aligned}$$

$$\begin{aligned} J^*_{T-j} - L_j(x_{oj}, u_{oj}) - J^*_{o,T-(j+1)} = &\min_{u_j} E\{L_x\delta x_j + \frac{1}{2}\delta x_j' L_{xx}\delta x_j + \delta x_j' L_{xu}\delta u_j \\ &+ L_u\delta u_j + \frac{1}{2}\delta u_j' L_{uu}\delta u_j + \Delta J^*_{T-(j+1)}\} \end{aligned}$$

Use $L_j(x_{oj}, u_{oj}) + J^*_{o,T-(j+1)} = J^*_{o,T-j}$ and assumption (23)

$$\begin{aligned} J^*_{T-j} - J^*_{o,T-j} = &\min_{u_j} E\{L_x\delta x_j + \frac{1}{2}\delta x_j' L_{xx}\delta x_j + \delta x_j' L_{xu}\delta u_j + L_u\delta u_j \\ &+ \frac{1}{2}\delta u_j' L_{uu}\delta u_j + g_{j+1} + E\{h^x_{j+1}\delta x_j + \frac{1}{2}\delta x_j' H_{j+1}\delta x_j / P^{j+1} / P^j\}\} \end{aligned}$$

Taking expectations over P^{j+1} yields

$$\begin{aligned} \Delta J^*_{T-j} := J^*_{T-j} - J^*_{o,T-j} = &\min_{u_j} E\{(L_x + h^x_{j+1})\delta x_j + \tfrac{1}{2}\delta x_j'(L_{xx} + H_{j+1})\delta x_j \\ &+ \delta x_j' L_{xu}\delta u_j + L_u\delta u_j + \tfrac{1}{2}\delta u_j' L_{uu}\delta u_j + g_{j+1}/P^j\}. \end{aligned} \tag{24}$$

Put (22) into (24) and set

$$L_x + h^x_{j+1} := k^x_j, \quad L_{xx} + H_{j+1} := K_j, \quad H_{T+1} = 0. \tag{25}$$

Then by taking expectations over P^j

$$\begin{aligned} \Delta J^*_{T-j} = &\min_{u_j} E\{k^x_j[(I - F_{x_j})^{-1}F_{x_{j-1}}\delta x_{j-1} + (I - F_{x_j})^{-1}F_u\delta u_j + (I - F_{x_j})^{-1}\varepsilon_j] \\ &+ \frac{1}{2}[(I - F_{x_j})^{-1}F_{x_{j-1}}\delta x_{j-1} + (I - F_{x_j})^{-1}F_u\delta u_j + (I - F_{x_j})^{-1}\varepsilon_j]' K_j \\ &\times [(I - F_{x_j})^{-1}F_{x_{j-1}}\delta x_{j-1} + (I - F_{x_j})^{-1}F_u\delta u_j + (I - F_{x_j})^{-1}\varepsilon_j] \\ &+ [(I - F_{x_j})^{-1}F_{x_{j-1}}\delta x_{j-1} + (I - F_{x_j})^{-1}F_u\delta u_j + (I - F_{x_j})^{-1}\varepsilon_j]' L_{xu}\delta u_j \\ &+ L_u\delta u_j + \frac{1}{2}\delta u_j' L_{uu}\delta u_j + g_{j+1}/P^{j-1}\} \end{aligned}$$

$$\begin{aligned}\Delta J^*_{T-j} = \min_{u_j} E\{ & k^x_j(I-F_{x_j})^{-1}F_{x_{j-1}}\delta x_{j-1} + [k^x_j(I-F_{x_j})^{-1}F_u + L_u]\delta u_j \\ & + \frac{1}{2}\delta x'_{j-1}[F'_{x_{j-1}}((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1}F_{x_{j-1}}]\delta x_{j-1} \\ & + \delta x'_{j-1}[F'_{x_{j-1}}((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1}F_u + F'_{x_{j-1}}((I-F_{x_j})^{-1})'L_{xu}]\delta u_j \\ & + \frac{1}{2}\delta u'_j[F'_u((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1})F_u + 2F'_u((I-F_{x_j})^{-1})'L_{xu} + L_{uu}]\delta u_j \\ & + \frac{1}{2}\varepsilon'_j((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1}\varepsilon_j + g_{j+1}/P^{j-1}\}.\end{aligned}$$

The terms which are higher than second order are dropped here.
Then use the following notations:

$$\begin{aligned} & \lambda^x := k^x_j[(I-F_{x_j})^{-1}F_{x_{j-1}} \\ & \lambda^u := k^x_j[(I-F_{x_j})^{-1}F_u + L_u \\ & \Lambda^{xx} := F'_{x_{j-1}}((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1}F_{x_{j-1}} \\ & \Lambda^{xu} := F'_{x_{j-1}}((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1}F_u + F'_{x_{j-1}}((I-F_{x_j})^{-1})'L_{xu} \\ & \Lambda^{uu} := F'_u((I-F_{x_j})^{-1})'K_j(I-F_{x_j})^{-1})F_u + 2F'_u((I-F_{x_j})^{-1})'L_{xu} + Luu \\ & \xi_j := (I-F_{x_j})^{-1}\varepsilon_j \end{aligned} \tag{26}$$

and get

$$\begin{aligned}\Delta J^*_{T-j} = \min_{u_j} E\{ & \lambda^x\delta x_{j-1} + \lambda^u\delta u_j + \tfrac{1}{2}\delta x'_{j-1}\Lambda^{xx}\delta x_{j-1} + \delta x'_{j-1}\Lambda^{xu}\delta u_j \\ & + \tfrac{1}{2}\delta u'_j\Lambda^{uu}\delta u_j + \tfrac{1}{2}\xi'_jK_j\xi_j + g_{j+1}/P^{j-1}\}.\end{aligned} \tag{27}$$

Apply the rule $E(x'Ax) = \hat{x}'A\hat{x} + tr(A\Sigma^{xx})$ and take expectations over P^{j-1}:

$$\begin{aligned}\Delta J^*_{T-j} = \min_{u_j} E\{ & \lambda^x\delta\hat{x}_{j-1/j-1} + \lambda^u\delta u_j + g_{j+1} + \tfrac{1}{2}\delta\hat{x}'_{j-1/j-1}\Lambda^{xx}\delta\hat{x}_{j-1/j-1} \\ & + \delta\hat{x}'_{j-1/j-1}\Lambda^{xu}\delta u_j + \tfrac{1}{2}\delta u'_j\Lambda^{uu}\delta u_j + \tfrac{1}{2}tr(\Lambda^{xx}\Sigma^{xx}_{j/j-1}) + \tfrac{1}{2}tr(K_j\Sigma^{\xi\xi})\},\end{aligned} \tag{28}$$

where $E(\xi) = 0$.
Next, minimize the objective function ΔJ^*_{T-j} with respect to δu_j:

$$\begin{aligned} \frac{\Delta J^*_{T-j}}{\delta u_j} &= \lambda^u + \delta\hat{x}'_{j-1/j-1}\Lambda^{xu} + \delta u'_j\Lambda^{uu} = 0 \\ \delta u_j &= -(\Lambda^{uu})^{-1}(\lambda^u + \Lambda^{xu}\delta\hat{x}_{j-1/j-1}) \end{aligned} \tag{29}$$

Put rule (29) into (27):

$$\begin{aligned}\Delta J^*_{T-j} = E\{ & \lambda^x\delta x_{j-1} + \lambda^u[-(\Lambda^{uu})^{-1}(\lambda^u + \Lambda^{xu}\delta\hat{x}_{j-1/j-1})] + \tfrac{1}{2}\delta x'_{j-1}\Lambda^{xx}\delta x_{j-1} \\ & + \delta x'_{j-1}\Lambda^{xu}[-(\Lambda^{uu})^{-1}(\lambda^u + \Lambda^{xu}\delta\hat{x}_{j-1/j-1})] + \tfrac{1}{2}[-(\Lambda^{uu})^{-1}(\lambda^u + \Lambda^{xu}\delta\hat{x}_{j-1/j-1})]'\Lambda^{uu} \\ & \times[-(\Lambda^{uu})^{-1}(\lambda^u + \Lambda^{xu}\delta\hat{x}_{j-1/j-1})] + \tfrac{1}{2}\xi'K_j\xi + g_{j+1}\}\end{aligned}$$

$$\begin{aligned}\Delta J^*_{T-j} = E\{ & g_{j+1} + (\lambda^x + \Lambda^{xu}(\Lambda^{uu})^{-1}\lambda^u)\delta x_{j-1} - \tfrac{1}{2}\lambda^u(\Lambda^{uu})^{-1}\lambda^u + \tfrac{1}{2}\delta x'_{j-1}\Lambda^{xx}\delta x_{j-1} \\ & + \tfrac{1}{2}\hat{x}'_{j-1/j-1}\Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux}\hat{x}_{j-1/j-1} - \delta x'_{j-1}\Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux}\delta\hat{x}_{j-1/j-1} + \tfrac{1}{2}\xi'K_j\xi\}\end{aligned}$$

Define

$$\Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux} := \Omega_{xx}. \tag{30}$$

Note that $\frac{1}{2}\xi' K_j \xi = \frac{1}{2}[E(\xi)K_j E(\xi) + tr(K_j \Sigma^{\xi\xi})] = \frac{1}{2}tr(K_j\Sigma^{\xi\xi})$ because $E(\xi) = 0$. Moreover $E\{\delta x'_{j-1}\Omega_{xx}\delta\hat{x}_{j-1/j-1}\} = \delta\hat{x}_{j-1/j-1}\Omega_{xx}\delta\hat{x}_{j-1/j-1}$. Thus,

$$\Delta J^*_{T-j} = g_{j+1} - \tfrac{1}{2}\lambda^u(\Lambda^{uu})^{-1}\lambda^u + \tfrac{1}{2}tr(K_j\Sigma^{\xi\xi}) + E\{(\lambda^x + \Lambda^{xu}(\Lambda^{uu})^{-1}\lambda^u)\delta x_{j-1}$$
$$+\tfrac{1}{2}\delta x'_{j-1}\Lambda^{xx}\delta x_{j-1}\} - \tfrac{1}{2}\delta\hat{x}_{j-1/j-1}\Omega_{xx}\delta\hat{x}_{j-1/j-1}.$$

Use the following rule:

$$\begin{aligned} &E\{\delta x_{j-1}\Omega_{xx}\delta x_{j-1}\} = \hat{x}_{j-1/j-1}\Omega_{xx}\delta\hat{x}_{j-1/j-1} + tr(\Omega_{xx}\Sigma^{xx}) \\ &\Rightarrow \quad \hat{x}_{j-1/j-1}\Omega_{xx}\delta\hat{x}_{j-1/j-1} = E\{\delta x_{j-1}\Omega_{xx}\delta x_{j-1}\} - tr(\Omega_{xx}\Sigma^{xx}) \end{aligned}$$

and get

$$\begin{aligned} \Delta J^*_{T-j} &= g_{j+1} - \tfrac{1}{2}\lambda^u(\Lambda^{uu})^{-1}\lambda^u + \tfrac{1}{2}tr(K_j\Sigma^{\xi\xi}) + E\{(\lambda^x + \Lambda^{xu}(\Lambda^{uu})^{-1}\lambda^u)\delta x_{j-1} \\ &\quad +\tfrac{1}{2}\delta x'_{j-1}\Lambda^{xx}\delta x_{j-1}\} - \tfrac{1}{2}(E\{\delta x_{j-1}\Omega_{xx}\delta x_{j-1}\} - tr(\Omega_{xx}\Sigma^{xx})) \\ &= g_{j+1} - \tfrac{1}{2}\lambda^u(\Lambda^{uu})^{-1}\lambda^u + \tfrac{1}{2}tr(K_j\Sigma^{\xi\xi}) + tr(\Omega_{xx}\Sigma^{xx}) \\ &\quad +E\{(\lambda^x + \Lambda^{xu}(\Lambda^{uu})^{-1}\lambda^u)\delta x_{j-1} + \tfrac{1}{2}\delta x'_{j-1}(\Lambda^{xx} - \Omega_{xx})\delta x_{j-1}\}. \end{aligned}$$

Defining

$$h^x_j := \lambda^x + \Lambda^{xu}(\Lambda^{uu})^{-1}\lambda^u, \quad H_j := \Lambda^{xx} - \Omega_{xx} := \Lambda^{xx} - \Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux} \tag{31}$$

and

$$g_j = g_{j+1} - \frac{1}{2}\lambda^u_j(\Lambda^{uu})^{-1}\lambda^u_j + \frac{1}{2}tr(K_j\Sigma^{\xi\xi}_{j/j} + \Omega_{xx}\Sigma^{xx}_j),$$

we get

$$\Delta J^*_{T-j} = g_j + E\{h^x_j\delta x_{j-1} + \tfrac{1}{2}\delta x'_{j-1}H_j\delta x_{j-1}\}. \tag{32}$$

This shows that ΔJ^*_{T-j} is a quadratic function of δx_{j-1}.

QED

Next, using the results of [31] we find the solution of the g recursion:

$$g_j = g_{j+1} - \frac{1}{2}\lambda^u_j(\Lambda^{uu})^{-1}\lambda^u_j + \frac{1}{2}tr(K_j\Sigma^{\xi\xi}_{j/j} + \Omega_{xx}\Sigma^{xx}_j).$$

Let us define a deterministic term $M_j = \frac{1}{2}\lambda^u_j(\Lambda^{uu})^{-1}\lambda^u_j$ and a stochastic term $N_j = \frac{1}{2}tr(K_j\Sigma^{\xi\xi}_{j/j} + \Omega_{xx}\Sigma^{xx}_j)$, so that $g_j = g_{j+1} - M_j + N_j$. By working backwards from period T the following can be shown:

$$g_{T-j} = g_{T-j+1} - \textstyle\sum_{i=T-j}^{T} M_i + \sum_{s=T-j}^{T} N_i \quad \mathit{and} \quad g_{T+1} = 0. \tag{33}$$

Next, solve it partially for the stochastic term N_j. For that define $\gamma_j = \gamma_{j+1} - M_j$ with $\gamma_{T+1} = 0$.
Using again the backward recursion get the equation:

$$\gamma_{T-j} = \gamma_{T-j+1} - \sum_{i=T-j}^{T} M_i = -\sum_{i=T-j}^{T} M_i.$$

Put it into (33) and get

$$g_{T-j} = \gamma_{T-j} + \sum_{i=T-j}^{T} N_i$$

or in general

$$g_{j+1} = \gamma_{j+1} + \sum_{i=j+1}^{T} N_i.$$

Thus

$$g_{j+1} = \gamma_{j+1} + \frac{1}{2}\sum_{i=j+1}^{T} tr(K_i \Sigma^{\xi\xi}_{i-1/i-1} + \Omega_{xx}\Sigma^{xx}_{i-1}) \tag{34}$$

and

$$\gamma_j = \gamma_{j+1} - \frac{1}{2}\lambda^{u}_{j-1}(\Lambda^{uu})^{-1}\lambda^{u}_{j-1}.$$

Use the fact that $\Delta J^*_{T-(t+1)}$ and ΔJ^*_{t+1} are equivalent; thus put (34) into (32) for $t+1$:

$$\Delta J^*_{t+1} = \gamma_{t+1} + \frac{1}{2}\left(\sum_{i=t+1}^{T} tr(K_i\Sigma^{\xi\xi} + \Omega_{xx}\Sigma^{xx})\right) + E\{h^x_{t+1}\delta x_t + \frac{1}{2}\delta x_t^{'} H_{t+1}\delta x_t / P^t\}. \tag{35}$$

Then

$$\begin{aligned} J^*_{t+1}(x_t) = J^*_{o,t+1} + \Delta J^*_{t+1} = C_{o,t+1} + \gamma_{j+1} + \tfrac{1}{2}(\textstyle\sum_{i=t+1}^{T} tr(K_i\Sigma^{\xi\xi} + \Omega_{xx}\Sigma^{xx})) \\ + E\{h^x_{t+1}\delta x_t + \tfrac{1}{2}\delta x_t^{'} H_{t+1}\delta x_t / P^t\}, \end{aligned} \tag{36}$$

where $C_{o,t+1} = \sum_{i=t+1}^{T} L_{t+1}(x_{o,t+1}, u_{o,t+1})$.

$$\begin{aligned} J^*_t &= \min_{u_t} E_{t-1}\{L_t(x_t,u_t) + J^*_{t+1}(x_t)\} \\ &= \min_{u_t} E\{L_t(x_t,u_t) + C_{o,t+1} + \gamma_{t+1} + \tfrac{1}{2}(\textstyle\sum_{i=t+1}^{T} tr(K_i\Sigma^{\xi\xi} + \Omega_{xx}\Sigma^{xx})) \\ &\quad + E\{h^x_{t+1}\delta x_t + \tfrac{1}{2}\delta x_t^{'} H_{t+1}\delta x_t / P^t / P^{t-1}\}\} \end{aligned} \tag{37}$$

$E\{h^x_{t+1}\delta x_t / P^{t-1}\} = 0$ and

$E\{\frac{1}{2}\delta x_t^{'} H_{t+1}\delta x_t / P^{t-1}\} = \frac{1}{2}\delta\hat{x}^{'}_{t/t-1} H_{t+1}\delta\hat{x}_{t/t-1} + \frac{1}{2}tr(H_{t+1}\Sigma^{xx}_{t/t-1}) = \frac{1}{2}tr(H_{t+1}\Sigma^{xx}_{t/t-1})$.

Then

$$\begin{aligned} J^*_t =: J^*_{t,d} = \min_{u_t} E_{t-1}\{L_t(x_t,u_t) + C_{o,t+1} + \gamma_{t+1} \\ + \tfrac{1}{2}\left(\textstyle\sum_{i=t+1}^{T} tr(K_i\Sigma^{\xi\xi}_{i/i} + \Omega_{xx}\Sigma^{xx}_{i/i})\right) + \tfrac{1}{2}tr(H_{t+1}\Sigma^{xx}_{t/t-1})\}. \end{aligned} \tag{38}$$

This objective function can be split into three parts:
deterministic:

$$J_{D,t} = L_t(x_t,u_t) + C_{o,t+1} + \gamma_{t+1}, \tag{39}$$

cautionary:

$$J_{c,t} = \frac{1}{2}\left(\sum_{i=t}^{T} tr(K_i\Sigma^{\xi\xi}_{i/i})\right) + \frac{1}{2}tr(H_{t+1}\Sigma^{xx}_{t/t-1}), \tag{40}$$

probing:

$$J_{p,t} = \frac{1}{2}\sum_{i=t+1}^{T} tr(\Omega_{xx}\Sigma^{xx}_{i/i}). \tag{41}$$

In this way the optimal control problem is solved and the components of the objective function are found. However, these are not the final results. In the next step, the stochastic parameters have to be taken into account, i.e., the extension of the system has to be considered.

3.5. Extension of the System

Recall that the algorithm deals with a quadratic criterion function when the parameters of the system equations are unknown. One technique to incorporate the uncertain parameters is to treat the random parameters as additional state variables.

Thus, apply the equations of the objective function to the system

$$\Psi_t = \begin{bmatrix} x_t \\ \theta_t \end{bmatrix}.$$

Then

$$\begin{aligned} L_t(\Psi_t, u_t) &= \tfrac{1}{2}[(x_t - \tilde{x}_t)' \vdots 0] \begin{bmatrix} W^{xx} & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{bmatrix} (x_t - \tilde{x}_t) \\ \cdots \\ 0 \end{bmatrix} \\ &+ [(x_t - \tilde{x}_t)' \vdots 0] \begin{bmatrix} W^{xu} \\ \cdots \\ 0 \end{bmatrix} [u_t - \tilde{u}_t] + \tfrac{1}{2}[u_t - \tilde{u}_t]' W^{uu} [u_t - \tilde{u}_t] \end{aligned} \tag{42}$$

and

$$f_t(\Psi_t, u_t) = \begin{bmatrix} f^x(\Psi_t, u_t) \\ f^\theta(\Psi_t, u_t) \end{bmatrix} = \begin{bmatrix} f(x_t, x_{t-1}, u_t, \theta_t) \\ I\theta_{t-1} \end{bmatrix}. \tag{43}$$

In order to calculate the adjusted components of the objective function (especially (40) and (41)) we need the terms H and Ω.
First, we calculate the adapted $\Omega_{\Psi\Psi}$ using Definition (30):

$$\begin{aligned} \Omega_{\Psi\Psi} &= \Lambda^{\Psi u}(\Lambda^{uu})^{-1}\Lambda^{u\Psi} = [F_{\Psi_{t-1}}((I - F_{\Psi_t})^{-1})' K_t (I - F_{\Psi_t})^{-1} F_u \\ &+ F_{\Psi_{t-1}}((I - F_{\Psi_t})^{-1})' L_{\Psi u}](\Lambda^{uu})^{-1}[F_u((I - F_{\Psi_t})^{-1})' K_t (I - F_{\Psi_t})^{-1} F_{\Psi_{t-1}} \\ &+ L'_{\Psi u}((I - F_{\Psi_t})^{-1})' F_{\Psi_{t-1}}] \\ &= \left(\begin{bmatrix} (F^x_{x_{t-1}})' & F^\theta_{x_{t-1}} \\ (F^x_{\theta_{t-1}})' & F^\theta_{\theta_{t-1}} \end{bmatrix} \begin{bmatrix} ((I - F^x_{x_t})^{-1})' & 0 \\ ((I - F^x_{x_t})^{-1} F^x_{\theta_t})' & I \end{bmatrix} \begin{bmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{bmatrix} \right. \\ &\times \begin{bmatrix} ((I - F^x_{x_t})^{-1})' & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{bmatrix} \begin{bmatrix} F^x_u \\ F^\theta_u \end{bmatrix} + \begin{bmatrix} F^x_{x_{t-1}} & F^\theta_{x_{t-1}} \\ F^x_{\theta_{t-1}} & F^\theta_{\theta_{t-1}} \end{bmatrix} \\ &\times \left. \begin{bmatrix} (I - F^x_{x_t})^{-1} & 0 \\ (I - F^x_{x_t})^{-1} F^x_{\theta_t} & I \end{bmatrix} \begin{bmatrix} L_{xu} \\ 0 \end{bmatrix} \right) (\Lambda^{uu})^{-1} \left(\begin{bmatrix} F'_u & 0 \end{bmatrix} \begin{bmatrix} (I - F^x_{x_t})^{-1} & 0 \\ (I - F^x_{x_t})^{-1} F^x_{\theta_t} & I \end{bmatrix} \right. \\ &\times \begin{bmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{bmatrix} \begin{bmatrix} (I - F^x_{x_t})^{-1} & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{bmatrix} \begin{bmatrix} F^x_{x_{t-1}} & F^x_{\theta_{t-1}} \\ F^\theta_{x_{t-1}} & F^\theta_{\theta_{t-1}} \end{bmatrix} \\ &+ \left. \begin{bmatrix} L_{ux} & 0 \end{bmatrix} \begin{bmatrix} (I - F^x_{x_t})^{-1} & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{bmatrix} \begin{bmatrix} F^x_{x_{t-1}} & F^x_{\theta_{t-1}} \\ F^\theta_{x_{t-1}} & F^\theta_{\theta_{t-1}} \end{bmatrix} \right), \end{aligned} \tag{44}$$

where

$$\begin{aligned} I - F_{\Psi_t} &= \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} F^x_{x_t} & F^x_{\theta_t} \\ F^\theta_{x_t} & F^\theta_{\theta_t} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} F^x_{x_t} & F^x_{\theta_t} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} I - F^x_{x_t} & (-F^x_{\theta_t}) \\ 0 & I \end{pmatrix} \end{aligned} \tag{45}$$

and

$$(I - F_{\Psi_t})^{-1} = \begin{pmatrix} (I - F^x_{x_t})^{-1} & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{pmatrix}. \tag{46}$$

We know that $F^{\theta}_{x_{t-1}} = 0$, $F^x_{\theta_{t-1}} = 0$, $F^{\theta}_{\theta_{t-1}} = I$, and $L_{xu} = W^{xu}$. Then we get

$$\begin{aligned} \Omega_{\Psi\Psi} = &\left(\begin{bmatrix} (F^x_{x_{t-1}})'(I - F^x_{x_t})^{-1} & 0 \\ ((I - F^x_{x_t})^{-1} F^x_{\theta_t})' & I \end{bmatrix} \begin{bmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{bmatrix} \begin{bmatrix} (I - F^x_{x_t})^{-1} F^x_{u_t} \\ 0 \end{bmatrix} \right. \\ &\left. + \begin{bmatrix} (F^x_{x_{t-1}})'(I - F^x_{x_t})^{-1} & 0 \\ ((I - F^x_{x_t})^{-1} F^x_{\theta_t})' & I \end{bmatrix} \begin{bmatrix} W^{xu} \\ 0 \end{bmatrix} \right) (\Lambda^{uu})^{-1} \\ &\times \left(\begin{bmatrix} (F^x_{u_t})'(I - F^x_{x_t})^{-1} & 0 \end{bmatrix} \begin{bmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{bmatrix} \right. \\ &\times \begin{bmatrix} (I - F^x_{x_t})^{-1} F^x_{x_{t-1}} & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{bmatrix} \\ &\left. + \begin{bmatrix} W^{ux} & 0 \end{bmatrix} \begin{bmatrix} (I - F^x_{x_t})^{-1} F^x_{x_{t-1}} & (I - F^x_{x_t})^{-1} F^x_{\theta_t} \\ 0 & I \end{bmatrix} \right). \end{aligned} \tag{47}$$

Define

$$D = (I - F^x_{x_t})^{-1} F^x_{\theta_t}. \tag{48}$$

Then using $A = (I - F^x_{x_t})^{-1} F^x_{x_{t-1}}$ and $B = (I - F^x_{x_t})^{-1} F^x_{u_t}$ we get the following:

$$\begin{aligned} \Omega_{\Psi\Psi} = &\left(\begin{bmatrix} A' & 0 \\ D' & I \end{bmatrix} \begin{bmatrix} K^{xx} B \\ K^{\theta x} B \end{bmatrix} + \begin{bmatrix} A' W^{xu} \\ D' W^{xu} \end{bmatrix} \right) (\Lambda^{uu})^{-1} \\ &\times \left(\begin{bmatrix} B' K^{xx} & B K^{x\theta} \end{bmatrix} \begin{bmatrix} A & D \\ 0 & I \end{bmatrix} + \begin{bmatrix} W^{ux} A \\ W^{ux} D \end{bmatrix} \right) \\ = &\begin{bmatrix} (A' K^{xx} B + A' W^{xu}) \\ (D' K^{xx} B + K^{\theta x} B + D' W^{xu}) \end{bmatrix} (\Lambda^{uu})^{-1} \\ &\times \begin{bmatrix} (B' K^{xx} A + W^{ux} A) & (B' K^{xx} D + B' K^{x\theta} + W^{ux} D) \end{bmatrix} =: \begin{bmatrix} \psi_1 & \psi_2 \\ \psi_3 & \psi_4 \end{bmatrix}. \end{aligned} \tag{49}$$

Use the definition $\Lambda^{xu} = A' K^{xx} B + A' W^{xu}$ and get the following:

$$\begin{aligned} \psi_1 &= \Lambda^{xu} (\Lambda^{uu})^{-1} \Lambda^{ux}, \\ \psi_2 &= \Lambda^{xu} (\Lambda^{uu})^{-1} (B' K^{xx} D + B' K^{x\theta} + W^{ux} D), \\ \psi_3 &= (D' K^{xx} B + K^{\theta x} B + D' W^{xu}) (\Lambda^{uu})^{-1} \Lambda^{ux}, \\ \psi_4 &= (D' K^{xx} B + K^{\theta x} B + D' W^{xu}) (\Lambda^{uu})^{-1} (B' K^{xx} D + B' K^{x\theta} + W^{ux} D). \end{aligned}$$

For the calculation of H we need the term $\Lambda^{\Psi\Psi}$:

$$\begin{aligned} \Lambda^{\Psi\Psi} &= (F_{\Psi_{t-1}})'((I - F_{\Psi_t})^{-1})' K_t (I - F_{\Psi_t})^{-1} F_{\Psi_t} \\ &= \begin{pmatrix} A' & 0 \\ D' & I \end{pmatrix} \begin{pmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{pmatrix} \begin{pmatrix} A & D \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A' K^{xx} A & A'(K^{xx} D + K^{x\theta}) \\ (D' K^{xx} + K^{\theta x}) A & D'(K^{xx} D + K^{x\theta}) + (K^{\theta x} D + K^{\theta\theta}) \end{pmatrix}. \end{aligned} \tag{50}$$

Now we can calculate the term H for the extended system:

$$H^{\Psi\Psi} = \Lambda^{\Psi\Psi} - \Lambda^{\Psi u} (\Lambda^{uu})^{-1} \Lambda^{u\Psi} = \Lambda^{\Psi\Psi} - \Omega_{\Psi\Psi}$$

$$H_t^{\Psi\Psi} = \begin{pmatrix} H^{xx} & H^{x\theta} \\ H^{\theta x} & H^{\theta\theta} \end{pmatrix}_t \tag{51}$$

Thus,

$$\begin{aligned} H^{xx} &= \Lambda^{xx} - \Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux} \\ H^{x\theta} &= A'(K^{xx}D + K^{x\theta}) - \Lambda^{xu}(\Lambda^{uu})^{-1}(B'K^{xx}D + B'K^{x\theta} + W^{ux}D) \\ H^{\theta\theta} &= D'(K^{xx}D + K^{x\theta}) + (K^{\theta x}D + K^{\theta\theta}) \\ &\quad -(D'K^{xx}B + K^{\theta x}B + D'W^{xu})(\Lambda^{uu})^{-1}(B'K^{xx}D + B'K^{x\theta} + W^{ux}D). \end{aligned} \tag{52}$$

After that we go back to the components of the objective function. According to (41) we need to calculate the term $\Omega_{\Psi\Psi}\Sigma_{t/t}^{\Psi\Psi}$.

$$\begin{aligned} \Omega_{\Psi\Psi}\Sigma_{t/t}^{\Psi\Psi} &= \begin{pmatrix} \psi_1 & \psi_2 \\ \psi_3 & \psi_4 \end{pmatrix}\begin{pmatrix} \Sigma^{xx} & \Sigma^{x\theta} \\ \Sigma^{\theta x} & \Sigma^{\theta\theta} \end{pmatrix}_{t/t} \\ &= \begin{pmatrix} \psi_1\Sigma_{t/t}^{xx} + \psi_2\Sigma_{t/t}^{\theta x} & \psi_1\Sigma_{t/t}^{x\theta} + \psi_2\Sigma_{t/t}^{\theta\theta} \\ \psi_3\Sigma_{t/t}^{xx} + \psi_4\Sigma_{t/t}^{\theta x} & \psi_3\Sigma_{t/t}^{x\theta} + \psi_4\Sigma_{t/t}^{\theta\theta} \end{pmatrix}. \end{aligned} \tag{53}$$

Thus,

$$\begin{aligned} J_{p,t} = \tfrac{1}{2}\textstyle\sum_{j=t+1}^{T} tr(\Omega_{\Psi\Psi}\Sigma_{j/j}^{\Psi\Psi}) &= \tfrac{1}{2}(\textstyle\sum_{j=t+1}^{T} tr[\Lambda^{xu}(\Lambda^{uu})^{-1}\Lambda^{ux}\Sigma_{j/j}^{xx}]) \\ &+tr[\Lambda^{xu}(\Lambda^{uu})^{-1}(B'(K^{xx}\overline{D} + K^{x\theta}) + W^{ux}\overline{D})\Sigma_{j/j}^{\theta x}] \\ &+\tfrac{1}{2}tr[((K^{\theta x} + (D)'K^{xx})B + (D)'W^{xu})(\Lambda^{uu})^{-1} \\ &\times(B'(K^{xx}D + K^{x\theta}) + W^{ux}D)\Sigma_{j/j}^{\theta\theta}]. \end{aligned} \tag{54}$$

Next we calculate $J_{C,t}$. According to (40) we need to calculate $tr(K_j^{\Psi\Psi}\Sigma_{j/j-1}^{\xi\xi})$ and $tr(H_{t+1}^{\Psi\Psi}\Sigma_{t/t-1}^{xx})$:

$$tr(K_j^{\Psi\Psi}\Sigma_{j/j}^{\xi\xi}) = tr\left[\begin{pmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{pmatrix}_j \begin{pmatrix} \Sigma^{\xi\xi} & 0 \\ 0 & 0 \end{pmatrix}_{j/j}\right] = tr(K_j^{xx}\Sigma_{j/j}^{\xi\xi}) \tag{55}$$

$$\begin{aligned} tr(H_{t+1}^{\Psi\Psi}\Sigma_{t/t-1}^{xx}) &= tr\left[\begin{pmatrix} H^{xx} & H^{x\theta} \\ H^{\theta x} & H^{\theta\theta} \end{pmatrix}_{t+1}\begin{pmatrix} \Sigma^{xx}\Sigma^{x\theta} \\ \Sigma^{\theta x}\Sigma^{\theta\theta} \end{pmatrix}_{t/t-1}\right] \\ &= tr(H_{t+1}^{xx}\Sigma_{t/t-1}^{xx} + H_{t+1}^{x\theta}\Sigma_{t/t-1}^{\theta x}) + tr(H_{t+1}^{\theta x}\Sigma_{t/t-1}^{x\theta} + H_{t+1}^{\theta\theta}\Sigma_{t/t-1}^{\theta\theta}) \\ &= tr(H_{t+1}^{xx}\Sigma_{t/t-1}^{xx}) + 2tr(H_{t+1}^{\theta x}\Sigma_{t/t-1}^{x\theta}) + tr(H_{t+1}^{\theta\theta}\Sigma_{t/t-1}^{\theta\theta}). \end{aligned} \tag{56}$$

Thus,

$$J_{c,t} = \frac{1}{2}[\sum_{i=j}^{T} tr(K_i^{xx}\Sigma_{i/i}^{\xi\xi})] + \frac{1}{2}tr(H_{t+1}^{xx}\Sigma_{t/t-1}^{xx}) + tr(H_{t+1}^{\theta x}\Sigma_{t/t-1}^{x\theta}) + \frac{1}{2}tr(H_{t+1}^{\theta\theta}\Sigma_{t/t-1}^{\theta\theta}). \tag{57}$$

Under the assumption $\gamma = 0$ (Kendrick (1981), Appendix J) the deterministic term is

$$\begin{aligned} J_{D,t} &= L_t(\Psi_t, u_t) + C_{o,t+1} + \gamma_{j+1} = L_t(x_t, u_t) + \textstyle\sum_{j=t+1}^{T}(L_j(x_{oj}, u_{oj})) \\ &= \tfrac{1}{2}(x_t - \widetilde{x}_t)'W^{xx}(x_t - \widetilde{x}_t) + (x_t - \widetilde{x}_t)'W^{xu}(u_t - \widetilde{u}_t) \\ &+ \tfrac{1}{2}(u_t - \widetilde{u}_t)'W^{uu}(u_t - \widetilde{u}_t) + \textstyle\sum_{j=t+1}^{T}[(x_{ot} - \widetilde{x}_t)'W^{xx}(x_{ot} - \widetilde{x}_t) \\ &+ (x_{ot} - \widetilde{x}_t)'W^{xu}(u_{ot} - \widetilde{u}_t) + \tfrac{1}{2}(u_{ot} - \widetilde{u}_t)'W^{uu}(u_{ot} - \widetilde{u}_t)], \end{aligned} \tag{58}$$

where $(x_{ot}, u_{ot}) = (x_t^*, u_t^*)$ is a nominal path.

In order to calculate the components of the objective function we need to determine one more term, $K_t^{\Psi\Psi}$.
Use $K_t^{\Psi\Psi} = L_{\Psi\Psi} + H_{t+1}^{\Psi\Psi}$

$$\begin{aligned} K_t^{\Psi\Psi} &= \begin{pmatrix} K^{xx} & K^{x\theta} \\ K^{\theta x} & K^{\theta\theta} \end{pmatrix}_t = \begin{pmatrix} L_{xx} & L_{x\theta} \\ L_{\theta x} & L_{\theta\theta} \end{pmatrix}_t + \begin{pmatrix} H^{xx} & H^{x\theta} \\ H^{\theta x} & H^{\theta\theta} \end{pmatrix}_{t+1} \\ &= \begin{pmatrix} W^{xx} & 0 \\ 0 & 0 \end{pmatrix}_t + \begin{pmatrix} H^{xx} & H^{x\theta} \\ H^{\theta x} & H^{\theta\theta} \end{pmatrix}_{t+1}, \end{aligned}$$

$$\begin{aligned} K_t^{xx} &= W_t^{xx} + H_{t+1}^{xx}, \\ K_t^{x\theta} &= H_{t+1}^{x\theta}, \\ K_t^{\theta x} &= H_{t+1}^{\theta x}, \\ K_t^{\theta\theta} &= H_{t+1}^{\theta\theta}, \end{aligned} \tag{59}$$

where $H_{T+1}^{xx} = O_{n\times n}$, $H_{T+1}^{x\theta} = O_{n\times p}$, $H_{T+1}^{\theta x} = O_{p\times n}$, $H_{T+1}^{\theta\theta} = O_{p\times p}$.
Use the linearized system, i.e., apply:
$A = (I - F_{x_t}^x)^{-1}F_{x_{t-1}}^x$, $B = (I - F_{x_t}^x)^{-1}F_{u_t}^x$, $L_u = x'W^{xu} + u'W^{uu} + (w^u)'$
$L_{uu} = W_t^{uu}$ and $L_{xu} = W_t^{xu}$
to (26) and obtain variables as used in Step I-2B in Section 3.3:

$$\begin{aligned} \Lambda_t^{xx} &= (A_t)'K_t^{xx}A_t, \\ \Lambda_t^{ux} &= (B_t)'K_t^{xx}A_t + W_t^{ux}A_t, \\ \Lambda_t^{xu} &= (\Lambda_t^{ux})', \\ \Lambda_t^{uu} &= (B_t)'K_t^{xx}B_t + 2(B_t)'W_t^{xu} + W_t^{uu}, \end{aligned} \tag{60}$$

$$\begin{aligned} \lambda_t^x &= k_t^x A_t, \\ \lambda_t^u &= k_t^x B_t + x_t'W_t^{xu} + u_t'W_t^{uu} + (w^u)'. \end{aligned} \tag{61}$$

4. Computational Aspects

In this paper we concentrate on mathematical details of the OPTCON algorithm. However, it is important to mention the computational characteristics of this algorithm. In order to show computational time and accuracy we used two different models: the MacRae model and the ATOPT model. As it would go beyond the scope of the present paper, we do not discuss these models and their economic interpretation (see detailed descriptions in [18]) but provide computational details of applying different versions of the OPTCON algorithm. We compared four different optimization strategies: deterministic solution (det), stochastic open-loop solution (OL) as described in Section 3.1, stochastic passive learning solution (OLF) as described in Section 3.2, and stochastic active learning solution (AL) as described in Section 3.3. In the case of the active learning strategy, a grid search with 100 grid points for each control variable was applied (see parameter Π in Step-I-[2] in Section 3.3). All calculations were performed on a Windows 10 computer with 16 GB RAM.

The OPTCON algorithm is programmed in MATLAB (MATLAB R2020a); the MATLAB computer code is available on request.

The MacRae model is a simple linear model with one state and one control variable. One parameter is stochastic and the optimal control was calculated for two periods only. Table 1 summarizes the computational performance of the OPTCON algorithm for the MacRae model. In the case of a linear-quadratic deterministic optimal control framework convergence can be shown analytically. This is true for the MacRae model, at least for the deterministic and the OL case. Moreover, as a check for the special case of a linear system, we compared the results of OPTCON with those of the same model calculated by available software and found them to be the same apart from round-off errors.

Table 1. Computational aspects of optimal control for the MacRae model

Method	Computational Time	Convergence [1]
det	0.024879 sec.	2
OL	0.036867 sec.	2
OLF	0.134725 sec.	2
AL	0.771425 sec.	2

[1] The number of iterations for convergence is available for the deterministic and OL solution techniques. In the case of the OLF and AL solution techniques the number of iterations for $S = 1$ is given, as this is the optimization for the whole planing horizon.

The ATOPT model is a small nonlinear model consisting of three equations, one control variable, and two stochastic parameters. The optimal control is calculated for five periods. Table 1 summarizes the computational performance of the OPTCON algorithm for the MacRae model. Table 2 summarizes the computational performance of the OPTCON algorithm for the ATOPT model.

Table 2. Computational aspects of optimal control for the ATOPT model

Method	Computational Time	Convergence
det	0.148106 sec.	15
OL	0.134021 sec.	7
OLF	0.265504 sec.	7
AL	15.726499 sec.	7

5. Concluding Remarks

In this paper, we described the OPTCON algorithm for the approximately optimal control of stochastic processes under a quadratic objective function. For systems with complete information, either deterministic or stochastic with known statistical characteristics of the disturbances, the open-loop version of OPTCON1 is suitable. OPTCON2 assumes partial information, in particular uncertain parameters of the system, with passive storage of information accruing during the control horizon, i.e., not used for control purposes. If active storage of information is possible, the dual control policy of OPTCON3 is appropriate. All three versions of OPTCON were programmed first in C# and then in MATLAB. We have applied OPTCON1 to various economic policy problems ([32,33], among others); Monte Carlo experiments with OPTCON2 and OPTCON3 have also yielded satisfactory results in terms of computing time. The results of approximately optimal policies were also as expected from the point of view of the economic problems under consideration. Due to the numerical nature of the algorithm, we cannot prove convergence in general; however, no problems of non-convergence have occurred so far. For future research, more applications are desirable to gain more insight into the different policies resulting from the various assumptions about the information structure of adaptive stochastic policy problems.

Author Contributions: conceptualization, R.N., D.B. and V.B.-N.; methodology, V.B.-N., R.N. and D.B.; software, D.B.; formal analysis, V.B.-N.; writing—original draft preparation, V.B.-N., D.B. and R.N.; writing—review and editing, R.N., D.B. and V.B.-N.; visualization, D.B. and V.B.-N.; supervision, R.N.; funding acquisition, V.B.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Science Fund FWF (grant T 1012-GBL).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Athans, M. The role and use of the stochastic linear-quadratic-Gaussian problem in control system design. *IEEE Trans. Autom. Control* **1971**, *16*, 529–552. [CrossRef]
2. Bellman, R. *Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 1957.
3. Bellman, R. *Adaptive Control Processes: A Guided Tour*; Princeton University Press: Princeton, NJ, USA, 1961.
4. Fel'dbaum, A.A. *Optimal Control Systems*; New York Academic Press: New York, NY, USA, 1965.
5. Aoki, M. *Optimization of Stochastic Systems: Topics in Discrete-Time Systems*; Academic Press: Cambridge, MA, USA, 1967; Volume 32.
6. Bertsekas, D.P. *Dynamic Programming and Optimal Control*; Athena Scientific: Belmont, MA, USA, 2005; Volume 1, 2.
7. Powell, W. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
8. Si, J.; Barto, A.; Powell, W.; Wunsch, D. *Handbook of Learning and Approximate Dynamic Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 2.
9. Matulka, J.; Neck, R. OPTCON: An Algorithm for the Optimal Control of Nonlinear Stochastic Models. *Ann. Oper. Res.* **1992**, *37*, 375–401. [CrossRef]
10. Blueschke-Nikolaeva, V.; Blueschke, D.; Neck, R. Optimal Control of Nonlinear Dynamic Econometric Models: An Algorithm and an Application. *Comput. Stat. Data Anal.* **2012**, *56*, 3230–3240. [CrossRef]
11. Kendrick, D.A.; Amman, H.M. A Classification System for Economic Stochastic Control Models. *Comput. Econ.* **2006**, *27*, 453–481. [CrossRef]
12. Bar-Shalom, Y.; Tse, E. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Trans. Autom. Control* **1974**, *19*, 494–500. [CrossRef]
13. Bar-Shalom, Y.; Tse, E. Caution, probing, and the value of information in the control of uncertain systems. In *Annals of Economic and Social Measurement*; NBER: Cambridge, MA, USA, 1976; Volume 5, pp. 323–337.
14. Tse, E.; Bar-Shalom, Y.; Meier, L. Wide-sense adaptive dual control for nonlinear stochastic systems. *IEEE Trans. Autom. Control* **1973**, *18*, 98–108. [CrossRef]
15. Kendrick, D.A. *Stochastic Control for Economic Models*; McGraw-Hill: New York, NY, USA, 1981.
16. Kendrick, D.A. Caution and probing in a macroeconomic model. *J. Econ. Dyn. Control* **1982**, *4*, 149–170. [CrossRef]
17. Amman, H.; Kendrick, D. Active learning monte carlo results. *J. Econ. Dyn. Control* **1994**, *18*, 119–124. [CrossRef]
18. Blueschke-Nikolaeva, V.; Blueschke, D.; Neck, R. OPTCON3: An Active Learning Control Algorithm for Nonlinear Quadratic Stochastic Problems. *Comput. Econ.* **2019**, *56*, 145–162. [CrossRef] [PubMed]
19. Behrens, D.; Neck, R. Approximating Solutions for Nonlinear Dynamic Tracking Games. *Comput. Econ.* **2015**, *45*, 407–433. [CrossRef]
20. Athans, M.; Falb, P. *Optimal Control: An Introduction to the Theory and Its Applications*; McGraw-Hill Book Co.: New York, NY, USA, 1966.
21. Chow, G.C. *Analysis and Control of Dynamic Economic Systems*; John Wiley & Sons: New York, NY, USA, 1975.
22. Chow, G.C. *Econometric Analysis by Control Methods*; John Wiley & Sons: New York, NY, USA, 1981.
23. Blueschke-Nikolaeva, V. *OPTCON2: An Algorithm for the Optimal Control of Nonlinear Stochastic Models*; Südwestdeutscher Verlag für Hochschulschriften: Saarbrücken, Germany, 2013.
24. Conn, A.R.; Gould, N.I.M.; Toint, P.L. *Trust-Region Methods*; Society for Industrial and Applied Mathematics and Mathematical Programming Society: Philadelphia, PA, USA, 2000.
25. Coleman, T.; Li, Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* **1996**, *6*, 418–445. [CrossRef]
26. Kelley, C. *Solving Nonlinear Equations with Newton's Method*; SIAM: Philadelphia, PA, USA, 2003.
27. Rheinboldt, W. *Methods for Solving Systems of Nonlinear Equations*; SIAM: Philadelphia, PA, USA, 1998.
28. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82D*, 33–45. [CrossRef]
29. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; Technical Report; Department of Computer Science, University of North Carolina: Chapel Hill, NC, USA, 2006.
30. Blueschke, D.; Blueschke-Nikolaeva, V.; Neck, R. Stochastic control of linear and nonlinear econometric models: Some computational aspects. *Comput. Econ.* **2013**, *42*, 107–118. [CrossRef]
31. Bar-Shalom, Y.; Tse, E.; Larson, R. Some recent advances in the development of closed-loop stochastic control and resource allocation algorithms. In Proceedings of the IFAC Symposium on Stochastic Control, Budapest, Hungary, 25–27 September 1974.

32. Neck, R.; Matulka, J. Stochastic optimum control of macroeconometric models using the algorithm OPTCON. *Eur. J. Oper. Res.* **1994**, *73*, 384–405. [CrossRef]
33. Neck, R.; Karbuz, S. Optimal budgetary and monetary policies under uncertainty: A stochastic control approach. *Ann. Oper. Res.* **1995**, *58*, 379–402. [CrossRef]