

VALUE AND POLICY FUNCTION APPROXIMATIONS IN INFINITE-HORIZON OPTIMIZATION PROBLEMS

GIORGIO GNECCO^{1, 2} AND MARCELLO SANGUINETI¹

¹ DEPARTMENT OF COMMUNICATIONS, COMPUTER, AND SYSTEM SCIENCES (DIST)
UNIVERSITY OF GENOA – VIA OPERA PIA 13, 16145 GENOVA – ITALY

² DEPARTMENT OF MATHEMATICS (DIMA)
UNIVERSITY OF GENOA – VIA DODECANESO 35, 16146 GENOVA – ITALY

E-MAIL: GIORGIO.GNECCO@DIST.UNIGE.IT, MARCELLO@DIST.UNIGE.IT

ABSTRACT. Suboptimal solutions to infinite-horizon dynamic optimization problems with continuous state are considered. An underlying dynamical system determining the state transition between each stage and the next one is modelled via the constraints $(x_t, x_{t+1}) \in D$, $t = 0, 1, \dots$, where X is the set to which the state vector belongs and $D \subseteq X \times X$ is a correspondence. An error analysis is performed for two cases: approximation of the value function and approximation of the optimal policy function. Structural properties of the dynamic optimization problems are derived, allowing to restrict a priori the approximation to families of functions characterized by certain smoothness properties. The two approximation approaches are compared and the respective pros and cons are highlighted.

AMS Classification: 90C39, 46N10

Keywords: Dynamic optimization, Dynamic programming, Infinite horizon, Approximation schemes, Suboptimal solutions.

The authors were partially supported by a PRIN Grant of the Italian Ministry for University and Research (Project “Models and Algorithms for Robust Network Optimization”).

JOURNAL OF DYNAMICAL SYSTEMS & GEOMETRIC THEORIES
VOL. 6, NUMBER 2 (2008) 123-147.
©TARU PUBLICATIONS

1. INTRODUCTION

We consider the following *infinite-horizon dynamic optimization problem*.

$$\begin{aligned} \text{Problem } \Sigma_\infty : \quad J^o(x_0) &= \sup \left\{ \sum_{t=0}^{\infty} \beta^t h_t(x_t, x_{t+1}) \right\} \\ \text{s.t.} \quad x_0 &\in X_0, (x_t, x_{t+1}) \in D_t, \quad t = 0, 1, \dots \end{aligned}$$

In this formulation, which follows [30], $X_t \subseteq \mathbb{R}^{d_t}$, called *state space*, is the set to which the vector x_t belongs; as each x_t is allowed to take a continuum of values in $\mathbb{R}^{d_t}$, Problem Σ_∞ is a *continuous-state* problem. At each time $t = 0, 1, \dots$, a *decision maker* has to choose the next state x_{t+1} , subject to the constraint $(x_t, x_{t+1}) \in D_t$, where $D_t \subseteq X_t \times X_{t+1}$ is a *correspondence* that models the transition from one stage to the following one. The functions $h_t : D_t \rightarrow \mathbb{R}$, $t = 0, 1, \dots$, model the *transition rewards*, which depend only on the current state and on the next one, chosen by the decision maker. The scalar $\beta \in (0, 1)$ is a fixed positive *discount factor*. Using an economic terminology, each $\beta^t h_t(x_t, x_{t+1})$ can be interpreted as the actualized value of the t -th transition reward. For a stochastic version of Problem Σ_∞ , see [40]. For simplicity of notations and without loss of generality, in the following we assume that d_t , X_t , h_t and D_t do not depend on the time t , so we write merely d , X , h and D , respectively. Problem Σ_∞ is *concave* when X and D are convex and the function h is concave on D .

Under mild hypotheses on the correspondence and the reward functions (and, in a stochastic framework, on the random variables), Problem Σ_∞ can be solved, in principle, by *Dynamic Programming* (DP) [5]. In some cases, a closed-form solution can be obtained by DP. Typically, however, closed-form solutions are not available, so one has to develop methodologies of approximate solution. Two widely-used approaches are value function approximation and policy function approximation [10, 34].

In the basic version of *value function approximation*, the current and next state vectors are discretized, then DP is applied to the simplified model (so, the DP equation has to be solved only for a finite number of state values), and finally an approximation is obtained for points not belonging to the discretization set.

Another way to solve Problem Σ_∞ consists in expressing the state at stage $t + 1$ as an unknown function of the state x_t at the previous stage and then searching

for a function, called *policy function* or simply *policy*, that does not violate the correspondence and maximizes the total reward. This approach is called *policy function approximation*. This entails a *functional optimization problem*, i.e., an optimization problem that is stated in an infinite-dimensional space [17]. Indeed, in this problem the unknown is a function belonging to a certain set of admissible functions, whose choice has to be guided by structural properties of the problem. The function's arguments are the components of the state vector. Since an analytical solution to this functional optimization task can be found only in particular cases and, even when it can be found, it may not be computable efficiently, one has to devise methodologies of approximate solution.

In this paper, first we investigate, for value and policy function approximations, the ways in which the error propagates from one stage to the next one. Then we derive structural properties of Problem Σ_∞ , allowing to restrict a priori the approximation to families of functions characterized by certain smoothness properties. Finally, we compare the two approximation approaches and point out the respective pros and cons.

The paper is organized as follows. Section 2 describes an instance of Problem Σ_∞ . Sections 3 and 4 investigate the error propagation in value and policy function approximation, respectively. Section 5 investigates smoothness properties of such functions. Section 6 compares value and policy function approximation and highlights the respective pros and cons. Section 7 considers possible extensions. To make the paper self-contained, we include an Appendix in which a technical result exploited in the paper is proven.

2. AN EXAMPLE OF PROBLEM Σ_∞

As we shall see later on, a basic assumption to obtain some of the results in this paper is the concavity of the transition reward. Although this might appear restrictive, it is quite common in applications. For the one-dimensional case, various examples of concave instances of Problem Σ_∞ can be found in [14]. In the following, we illustrate a multidimensional concave problem from [42], known as “multisector optimal growth model”, and we cast it in the framework of Problem Σ_∞ . Examples of applications to which our approach applies can be found in [21, 23].

Given a pure consumption good $y \in \mathbb{R}$ (where “pure” means that it is not used as an input in a production process), let $\mathcal{T}(x, p)$, where $x, p \in \mathbb{R}^d$, be the maximum amount of y that can be produced in a unit time interval from the capital goods $x^{(i)}, i = 1, \dots, d$, under the constraint that an amount $p^{(i)}$ of each capital good is produced during the same time interval. Under appropriate assumptions, it can be proven that $\mathcal{T}(x, p)$ is concave (but not strictly concave) [42]. Let $u(c)$ be a concave nondecreasing utility function associated with the consumption of an amount c of the pure consumption good. The target consists in maximizing

$$\begin{aligned} J^o(x_0) &= \sup \left\{ \sum_{t=0}^{\infty} \beta^t u(\mathcal{T}(x_t, p_t)) \right\} \\ \text{s.t. } \quad x_0 &\in X, \quad (x_t, p_t) \in \text{dom}(\mathcal{T}), \quad x_{t+1} = p_t + (1 - \mu)x_t, \quad t = 0, 1, \dots \end{aligned}$$

Here, $\text{dom}(\mathcal{T})$ is the domain of the function $\mathcal{T}$ (not all combinations of x_t and p_t are allowed, depending on the production functions that implicitly define $\mathcal{T}$ and its domain [42]), and $\mu \in [0, 1]$ is the rate of depreciation of the capital, which is assumed to be the same for all the x_i . By removing p_t through the constraints above, the utility per stage can be expressed as a function of x_t and x_{t+1} , thus obtaining

$$\begin{aligned} J^o(x_0) &= \sup \left\{ \sum_{t=0}^{\infty} \beta^t u(\mathcal{T}(x_t, x_{t+1} - (1 - \mu)x_t)) \right\} \\ \text{s.t. } \quad x_0 &\in X, \quad (x_t, x_{t+1}) \in D, \quad t = 0, 1, \dots, \end{aligned}$$

where the correspondence D is given by

$$D \triangleq \{(x, z) \in \mathbb{R}^{2d} : x, z \in \mathbb{R}^d \text{ and } (x, z - (1 - \mu)x) \in \text{dom}(\mathcal{T})\}.$$

By basic results on concavity-preserving operations [12, Section 3.2], it follows that $u(\mathcal{T}(x_t, x_{t+1} - (1 - \mu)x_t))$ is concave. So, the multisector optimal growth model is a concave instance of Problem Σ_{∞} .

3. ERROR PROPAGATION IN VALUE FUNCTIONS APPROXIMATION

Let the following assumption hold.

Assumption 3.1. *The transition reward h is bounded and continuous on D and for every $x \in X$ the set $D(x) \triangleq \{y : (x, y) \in D\}$ is nonempty.*

Let $\mathcal{B}(X)$ be the space of bounded functions on X and $T : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ Bellman's operator, defined for every $f \in \mathcal{B}(X)$ as [14, p. 176]

$$(Tf)(x) := \sup_{y \in D(x)} [h(x, y) + \beta f(y)].$$

In terms of Bellman's operator, the solution to Problem Σ_∞ satisfies the *stationary Bellman's equation* [14, Theorem 2.1, p. 38]

$$(1) \quad J^o(x) = (TJ^o)(x).$$

Endowed with the supremum norm, $\mathcal{B}(X)$ is a Banach space and for $\beta \in (0, 1)$ T is a *contraction operator* thereon [14, Corollary 9.1, p. 178]. In this case, J^o is the unique fixed point of T and can be determined by the iterated application of Bellman's operator, starting from an initial guess. This procedure is called the *Dynamic Programming* (DP) algorithm.

In practice, a solution in closed form to Problem Σ_∞ can be derived via DP only in particular cases. Usually, one must resort to approximation approaches. In the basic version of *Approximate Dynamic Programming* (ADP) [10, 34], the state space is discretized and DP is applied only in correspondence of a finite number of state values. For points in the state space not belonging to the discretization set, a suitable *value function approximation* is determined.

We shall need the following notations and definitions. For $\Omega \subseteq \mathbb{R}^d$, by $\mathcal{C}(\Omega)$ and $\|\cdot\|_{\infty(\Omega)}$ we denote the set of continuous functions on Ω and their supremum norm, respectively. The same notation is used for continuous vector-valued functions $g : \Omega \rightarrow \mathbb{R}^m$, for which

$$\|g\|_{\infty(\Omega)} \triangleq \sup_{\omega \in \Omega} \sqrt{\sum_{j=1}^m |g_j(\omega)|^2}.$$

For a positive integer s , $\mathcal{C}^s(\Omega)$ is the set of functions on Ω with partial derivatives that are continuous up to the order s . Given a Lebesgue-measurable set $\Omega \subseteq \mathbb{R}^d$ and $1 \leq p \leq \infty$, by $\mathcal{L}_p(\Omega)$ we denote the correspondent Lebesgue space, where

integration is performed with respect to the Lebesgue measure, and by $\|\cdot\|_{\mathcal{L}_p(\Omega)}$ the classical $\mathcal{L}_p(\Omega)$ -norm.

In order to perform an analysis of the approximation error, in particular to investigate how it propagates through the iterations of ADP, it is necessary to measure it by an appropriate norm. To this end, it is mandatory to take into account that at each step the original DP algorithm requires to perform one maximization. The presence of the maximization step makes $\mathcal{L}_p$ -norms with $p \neq \infty$ unsuitable for the error-propagation analysis. This is due to the fact that for $p \neq \infty$ and $f_1, f_2 \in \mathcal{B}(X)$, a small value of $\|f_1 - f_2\|_{\mathcal{L}_p(X)}$ does not necessarily imply a small value of $\|Tf_1 - Tf_2\|_{\mathcal{L}_p(X)}$ (unless X is bounded and additional smoothness assumptions on f_1 and f_2 are made). Instead, for the supremum norm one has the following estimate [14, Corollary 9.1]; for completeness, we give its proof.

Proposition 3.2. [14, Corollary 9.1] *If Assumption 3.1 holds and f_1, f_2 are bounded and continuous on X , then*

$$\|Tf_1 - Tf_2\|_{\infty(X)} \leq \beta \|f_1 - f_2\|_{\infty(X)}.$$

Proof. By the definition of Bellman's operator we get

$$\begin{aligned} & (Tf_1)(x) - (Tf_2)(x) \\ &= \sup_{y \in D(x)} [h(x, y) + \beta f_1(y)] - \sup_{y \in D(x)} [h(x, y) + \beta f_2(y)] \\ &\leq \sup_{y \in D(x)} [h(x, y) + \beta f_2(y)] + \beta \sup_{x \in X} |f_1(x) - f_2(x)| \\ &\quad - \sup_{y \in D(x)} [h(x, y) + \beta f_2(y)] \\ &= \beta \sup_{x \in X} |f_1(x) - f_2(x)| \end{aligned}$$

A similar upper bound on $(Tf_2)(x) - (Tf_1)(x)$ is obtained by reversing the roles of f_1 and f_2 . ■

The next proposition estimates the error propagation when a family

$$\mathcal{F}_J \subset \{f : X \rightarrow \mathbb{R} \text{ bounded and continuous}\}$$

of functions is used at each iteration of ADP (for simplicity, we assume that at all the iterations the same family $\mathcal{F}_J$ of approximators is used). In order to focus on

the approximation capabilities of the family of approximators $\mathcal{F}_J$, in the following analysis we neglect the effect of the finite number of samples of the functions to be approximated. Supposing that each iteration of ADP has ideally at its disposal an infinite number of samples allows us to use directly the supremum norm in the analysis¹.

Proposition 3.3. *Let Assumption 3.1 hold. Let $M > 0$ be such that $\|h\|_{\infty(D)} \leq M$, $\mathcal{F}_J$ a family of bounded and continuous functions on X , and $\varepsilon > 0$ such that there exists $f \in \mathcal{F}_J$ satisfying $\|J^o - f\|_{\infty(X)} \leq \varepsilon$. Then for every positive integer N , after N iterations of ADP starting with $\tilde{J}_N^o = 0$ and such that $\tilde{J}_t^o \in \mathcal{F}_J$, $t = N - 1, \dots, 0$, the worst-case approximation error of ADP satisfies*

$$\|J^o - \tilde{J}_0^o\|_{\infty(X)} \leq \frac{1 - (2\beta)^N}{1 - 2\beta} \varepsilon + (2\beta)^N \frac{M}{1 - \beta}.$$

Proof. J^o is bounded and continuous as it is the unique fixed point of Bellman's operator on the space $\mathcal{B}(X)$. By the formulation of Problem Σ_∞ and the hypothesis $\|h\|_{\infty(D)} \leq M$, we get $\|J^o\|_{\infty(X)} \leq \frac{M}{1-\beta}$. Let $\eta_N = \frac{M}{1-\beta}$. Using a backward induction argument, for $t = N - 1, \dots, 0$ assume that at the iteration $t + 1$ of ADP, $\tilde{J}_{t+1}^o \in \mathcal{F}_J$ is such that $\|J^o - \tilde{J}_{t+1}^o\|_{\infty(X)} \leq \eta_{t+1}$ for some $\eta_{t+1} > 0$. Let $\hat{J}_t^o = T\tilde{J}_{t+1}^o$; recalling that J^o is the unique fixed point of T , Proposition 3.2 gives

$$\|J^o - \hat{J}_t^o\|_{\infty(X)} = \|TJ^o - T\tilde{J}_{t+1}^o\|_{\infty(X)} \leq \beta \|J^o - \tilde{J}_{t+1}^o\|_{\infty(X)} \leq \beta \eta_{t+1}.$$

Before moving to the next iteration of ADP, one has to find an approximation $\tilde{J}_t^o \in \mathcal{F}_J$ for J^o . Such an approximation has to be obtained from $\hat{J}_t^o$ (which in general does not belong to $\mathcal{F}_J$), because J^o is unknown. On a neighborhood of radius $\beta\eta_{t+1}$ of $\hat{J}_t^o$, there may exist a function I_t , different from J^o , that can be approximated by $\tilde{J}_t^o \in \mathcal{F}_J$ with an error less than or equal to ε , too. So, $\|J^o - I_t\|_{\infty(X)} \leq \|J^o - \hat{J}_t^o\|_{\infty(X)} + \|\hat{J}_t^o - I_t\|_{\infty(X)} \leq 2\beta\eta_{t+1}$, from which we get

$$\|J^o - \tilde{J}_t^o\|_{\infty(X)} \leq 2\beta\eta_{t+1} + \varepsilon \triangleq \eta_t.$$

As $\eta_N = \frac{M}{1-\beta}$, after N iterations of ADP we obtain

¹A refined analysis which takes into account the effect of a finite number of samples may be done by using tools from Statistical Learning Theory [41] and exploiting structural properties of the functions to be approximated.

$$\begin{aligned}
\|J^o - \tilde{J}_0^o\|_{\infty(X)} \leq \eta_0 &= \varepsilon + 2\beta\eta_1 = \varepsilon + 2\beta\varepsilon + 4\beta^2\eta_2 = \dots = \varepsilon(N) \\
&= \sum_{t=0}^{N-1} (2\beta)^t \varepsilon + (2\beta)^N \frac{M}{1-\beta} \\
&= \frac{1 - (2\beta)^N}{1-2\beta} \varepsilon + (2\beta)^N \frac{M}{1-\beta}.
\end{aligned}$$

■

By straightforward computations, if $0 < \beta < \frac{1}{2}$ and $0 < \varepsilon < \frac{1-2\beta}{1-\beta}M$, then the optimal upper bound from Proposition 3.3 is obtained in the limit $N \rightarrow \infty$, for which one has $\varepsilon_\infty = \lim_{N \rightarrow \infty} \varepsilon(N) = \frac{\varepsilon}{1-2\beta}$.

4. ERROR PROPAGATION IN POLICY FUNCTIONS APPROXIMATION

An optimal policy function g^o for Problem Σ_∞ can be obtained as a byproduct of DP:

$$(2) \quad g^o(x) \in \operatorname{argmax}_{y \in D(x)} [h(x, y) + \beta J^o(y)].$$

In writing (2), we are supposing that for every $x \in X$, $\max_{y \in D(x)} [h(x, y) + \beta J^o(y)]$ exists. To investigate when this is the case, let us introduce the following assumption. Recall that for $\Omega \subseteq \mathbb{R}^d$ and $f : \Omega \rightarrow \mathbb{R}$, f is *upper semicontinuous* on Ω if for every $x \in \Omega$ and every sequence $\{x_i\}$ such that $x_i \in \Omega$ and $\lim_{i \rightarrow \infty} x_i = x$, one has $f(x) \geq \limsup_{i \rightarrow \infty} f(x_i)$.

Assumption 4.1. *The following hold:*

- (i) D is closed and for every $x \in X$ the set $D(x)$ is nonempty;
- (ii) h is bounded and upper semicontinuous on D ;
- (iii) there exist $c_1, c_2 \geq 0$ such that for every $(x, y) \in D$ one has $\|y\| \leq c_1 + c_2\|x\|$;
- (iv) $h(x_t, x_{t+1}) \leq L_{x_0, t}$, where $\sum_{t=0}^{\infty} L_{x_0, t} \beta^t < \infty$ and $\{x_t\}_{t=0}^{\infty}$ is any sequence of states such that $x_{t+1} \in D(x_t)$.

Under Assumption 4.1, there exists an optimal policy function with initial state x_0 [14, Proposition 4.7, p. 78]. The next assumption guarantees that, if Problem Σ_∞ admits an optimal policy function, then such a function is unique [14, Proposition 4.5, p. 76], so the symbol “ $\in$ ” in (2) can be replaced by equality.

Assumption 4.2. *Problem Σ_∞ is concave (i.e., X and D are convex and the function h is concave on D). Moreover, one of the following conditions holds:*

- (i) *for every $x \in X$, $h(x, \cdot)$ is strictly concave on $D(x)$;*
- (ii) *for every every $0 < \theta < 1$ and every $(x, y), (x', y') \in D$ such that $x \neq x'$, one has $h(\theta x + (1 - \theta)x', \theta y + (1 - \theta)y') > \theta h(x, y) + (1 - \theta)h(x', y')$.*

So, if both Assumptions 4.1 and 4.2 hold, then there exists a unique optimal policy function for Problem Σ_∞ .

The next proposition estimates the error propagation when a family

$$\mathcal{F}_g = \{f : X \rightarrow X \text{ bounded and continuous, } f(x) \in D(x) \forall x \in X\}$$

of functions is employed to solve Problem Σ_∞ by policy approximation. For any policy $f \in \mathcal{F}_g$, let

$$\begin{aligned} \tilde{J}_f^o(\tilde{x}_0) &= \left\{ \sum_{t=0}^{\infty} \beta^t h(\tilde{x}_t, \tilde{x}_{t+1}) \right\} \\ \text{s.t. } \tilde{x}_0 &\in X_0, \tilde{x}_{t+1} = f(\tilde{x}_t), \quad t = 0, 1, \dots \end{aligned}$$

Proposition 4.3. *Let Problem Σ_∞ admit an optimal policy g^o , h and g^o be bounded and Lipschitz continuous, with Lipschitz constants bounded from above by L_h and L_{g^o} , resp., and $M > 0$ such that $\|h\|_{\infty(D)} \leq M$. Let $\varepsilon > 0$ and $\mathcal{F}_g$ be a family of bounded and continuous functions such that there exists $f \in \mathcal{F}_g$ satisfying $\|g^o - f\|_{\infty(X)} \leq \varepsilon$. Then for every positive integer N*

$$\|J^o - \tilde{J}_f^o\|_{\infty(X)} \leq L_h \sum_{t=0}^{N-1} \beta^t \sqrt{\theta_t^2 + \theta_{t+1}^2} + \frac{\beta^N}{1 - \beta} M,$$

where $\theta_0 = 0$ and $\theta_{t+1} = L_{g^o} \theta_t + \varepsilon$, $t = 0, 1, \dots$

Proof. For any initial state $\tilde{x}_0 = x_0$, we compare the behavior of the two dynamical systems

$$x_{t+1} = g^o(x_t)$$

and

$$\tilde{x}_{t+1} = f(\tilde{x}_t),$$

where $f \in \mathcal{F}_g$ and $\|g^o - f\|_{\infty(X)} \leq \varepsilon$. By the Lipschitz continuity of h we get

$$(3) \quad \|J^o - \tilde{J}_f^o\|_{\infty(X)} \leq \sum_{t=0}^{N-1} L_h \beta^t \sqrt{\|x_t - \tilde{x}_t\|^2 + \|x_{t+1} - \tilde{x}_{t+1}\|^2} + \frac{\beta^N}{1-\beta} M,$$

where the series starting from $t = N$ has been bounded from above by $\sum_{t=N}^{\infty} \beta^t M = \frac{\beta^N}{1-\beta} M$.

The Lipschitz continuity of g^o and the triangle inequality give

$$\begin{aligned} \|x_{t+1} - \tilde{x}_{t+1}\| &= \|g^o(x_t) - f(\tilde{x}_t)\| = \|g^o(x_t) - g^o(\tilde{x}_t) + g^o(\tilde{x}_t) - f(\tilde{x}_t)\| \\ &\leq \|g^o(x_t) - g^o(\tilde{x}_t)\| + \|g^o(\tilde{x}_t) - f(\tilde{x}_t)\| \\ (4) \quad &\leq L_{g^o} \|x_t - \tilde{x}_t\| + \|g^o(\tilde{x}_t) - f(\tilde{x}_t)\|. \end{aligned}$$

Let $\tilde{\theta}_t \triangleq \|x_t - \tilde{x}_t\|$. Then $\tilde{\theta}_0 = 0$ and equation (4) provides $\tilde{\theta}_{t+1} \leq L_{g^o} \tilde{\theta}_t + \varepsilon$. The proof is completed by using (3) and noting that $\theta_t \geq \tilde{\theta}_t$ for each nonnegative integer t . ■

5. STRUCTURAL PROPERTIES OF THE VALUE AND OPTIMAL POLICY FUNCTIONS

Structural properties of the value and optimal policy functions are very useful in order to find a suitable approximating family. Some of these properties can be obtained proving that they are preserved by one iteration of the DP algorithm, then they can be extended to a finite and sometimes countable number of iterations [37].

5.1. A short survey of known structural properties. For certain families of problems, explicit solutions are available. The typical example is the LQ problem, i.e. when the correspondence D in Problem Σ_{∞} is defined through a linear dynamic system and the transition reward is quadratic. In this case, the value function J^o is quadratic [8, Chapter 4] and the optimal policy function g^o is linear in the state vector. For some reward functions, another simple case for which explicit solutions are available is Gale's cake-eating problem [18, 14].

- If the transition reward h satisfies a certain condition of strict concavity and at least one potentially optimal path [14, p. 41] exists for each initial state, then J^o is strictly concave [14, p. 73].
- Under some conditions the value function is monotone. More precisely, if $D(x)$ is monotonic and for every $y \in X$ the transition reward $h(\cdot, y)$ is nondecreasing, then J^o is nondecreasing, too [14, p. 77].

- Some conditions guarantee monotonicity of the optimal policy g^o . In [14, p. 85], it is proved that for one-dimensional systems, monotonicity of g^o is related to super-modularity of the transition reward. Moreover, under assumptions of interiority and continuity of the optimal policy, its local behaviour is related to the local super-modularity or sub-modularity of the transition reward [14, p. 86]. The relationship between super-modularity (or sub-modularity) and monotonicity of g^o holds not only for one-dimensional systems but also, in a stochastic setting, for Markov decision processes with a countable state space [35, Theorem 4.7.4].
- For a concave Problem Σ_∞ , it is possible to prove differentiability of J^o when the pre-image of a point under the optimal policy is an interior point and for every $y \in X$ the transition reward $h(\cdot, y)$ is differentiable in the first argument (Benveniste-Scheinkman's envelope theorem; see [6] and [14, p. 90]).
- Under strong concavity of the transition reward, differentiability of g^o was proven in [31]. Among other conditions, this result requires a condition of uniform interiority of g^o .
- In [14, p. 174] it is shown that twice local differentiability of J^o is related to local differentiability of g^o . More precisely, under the hypotheses of Benveniste-Scheinkman's envelope theorem, if g^o is locally differentiable, then twice local differentiability of J^o readily follows from the chain rule.
- Higher-order differentiability of g^o was obtained locally in [14, p. 159] by the analysis of the stability of the discrete time Euler-Lagrange equation [14, p. 114]. Among the hypotheses there is the existence of an interior fixed point of the optimal policy.
- In [30], Lipschitz continuity of g^o is proved under strong concavity hypotheses and an upper bound on the Lipschitz constant is given, too. For a special choice of the correspondence D , this result does not require interiority of g^o .

- For every policy function belonging to certain families of functions, there exists an instance of Problem Σ_∞ for which such a function is the optimal policy for that problem (see [14, p. 183] and [32]).²

5.2. A global smoothness result. For an open set $\Omega \subseteq \mathbb{R}^d$, a positive integer m , and $1 \leq p \leq \infty$, $\mathcal{W}_p^m(\Omega)$ is the *Sobolev space* of functions whose weak partial derivatives up to the order m are in $\mathcal{L}_p(\Omega)$. The d -dimensional *Fourier transform* is defined as the unique continuous operator from $\mathcal{L}_2(\mathbb{R}^d)$ to $\mathcal{L}_2(\mathbb{R}^d)$ such that, when restricted to $f \in \mathcal{L}_1(\mathbb{R}^d) \cap \mathcal{L}_2(\mathbb{R}^d)$, one has $f(x) \mapsto \hat{f}(s) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\langle x, s \rangle} f(x) dx$. For $r > 0$, the *Bessel potential* of order r on $\mathbb{R}^d$, denoted by $G_r(x)$, is the inverse Fourier transform of the function

$$\hat{G}_r(s) = (2\pi)^{-\frac{d}{2}} (1 + \|s\|^2)^{-r/2}.$$

For $r > 0$ and $1 \leq p \leq \infty$, $\mathcal{B}_p^r(\mathbb{R}^d)$ is the *Bessel potential space* whose elements are functions u such that $u = f * G_r$, where $f \in \mathcal{L}_p(\mathbb{R}^d)$ and $\|u\|_{\mathcal{B}_p^r(\mathbb{R}^d)} \triangleq \|f\|_{\mathcal{L}_p(\mathbb{R}^d)}$ [39, p. 134]. For $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$, $f * g$ is the *convolution* of f and g , i.e., $(f * g)(x) = \int_{\mathbb{R}^d} f(y)g(x-y)dy$. For a set $\Omega \subseteq \mathbb{R}^d$, we denote by $\text{int}(\Omega)$ its *topological interior*. If $\Omega_1 \subset \Omega \subseteq \mathbb{R}^d$ and f is a function on Ω , we denote by $f|_{\Omega_1}$ its restriction to Ω_1 (similarly, f is an extension to Ω of $f|_{\Omega_1}$).

We shall prove, for the value and optimal policy functions, global smoothness properties expressed via Bessel potentials. Norms in Bessel potential spaces correspond to a concept of smoothness defined in terms of smoothing operators of the form of convolutions with Bessel potentials and are related to Sobolev norms, which correspond to a concept of smoothness defined in terms of $\mathcal{L}_p$ norms of partial derivatives [39, Chapter 5]. Approximation properties of functions belonging to Bessel potential spaces are investigated, e.g., in [19, 22].

We use the following notations for partial derivatives. ∇ denotes the *gradient* when it is applied to a scalar function and the *Jacobian* when it is applied to a vector function. Let d_1 and d_2 be positive integers and $x \in \mathbb{R}^{d_1}$, $y \in \mathbb{R}^{d_2}$ independent variables. For a composition function $f(g(x, y), h(x, y))$, by $\nabla_1 f(g(x, y), h(x, y))$ we denote the vector of partial derivatives of f with respect to its first vector argument

²These are “quite large” families of functions and it can be shown that there exist instances of Problem Σ_∞ for which the dynamic system $x_{t+1} = g^\circ(x_t)$ obtained by iterating the optimal policy exhibits chaotic behaviour [14, p. 183] and [32].

g and by $\nabla_x f(g(x, y), h(x, y))$ the full gradient of f with respect to x . We use the notation ∇^2 for the *Hessian*. For example, the *Hessian* of $f(x, y)$ is

$$\nabla^2 f(x, y) = \begin{pmatrix} \nabla_{1,1}^2 f(x, y) & \nabla_{1,2}^2 f(x, y) \\ \nabla_{2,1}^2 f(x, y) & \nabla_{2,2}^2 f(x, y) \end{pmatrix}.$$

Assumption 5.1. *The following hold:*

- (i) $X \subset \mathbb{R}^d$ is compact, convex and has nonempty interior;
- (ii) $D \subseteq X \times X$ is compact, convex and has nonempty interior;
- (iii) for each $x \in X$, the set $D(x)$ is nonempty;
- (iv) h is continuous and concave.

Proposition 5.2. *Let Assumption 5.1 be satisfied, $D = X \times X$, $h \in \mathcal{C}^2(D)$, and $\nabla^2 h(x, y)$ be non-singular for every $(x, y) \in D$. Then the following hold.*

- (i) *The optimal policy function g^o is Lipschitz continuous on X and for every $1 < p < \infty$ there exists a function $\bar{g}^{o,p} \in \mathcal{B}_p^1(\mathbb{R}^d)$ such that $g^o = \bar{g}^{o,p}|_X$.*
- (ii) *The value function J^o is continuous and has Lipschitz continuous partial derivatives, and for every $1 \leq p < \infty$ there exists a function $\bar{J}^{o,p} \in \mathcal{B}_p^2(\mathbb{R}^d)$ such that $J^o = \bar{J}^{o,p}|_X$.*

Proof. (i) The Lipschitz continuity of the optimal policy function g^o follows by [30, Theorem 5.2], which also gives its uniqueness. Then by [25, Theorem 6.12, p. 46], g^o belongs to the Sobolev space $\mathcal{W}_\infty^1(\text{int}(X))$. As X is bounded, for every $1 \leq p < \infty$ the inclusion $\mathcal{W}_\infty^1(\text{int}(X)) \subset \mathcal{W}_p^1(\text{int}(X))$ holds, so $g^o \in \mathcal{W}_p^1(\text{int}(X))$.

Since X is bounded and convex, by Sobolev's extension theorem [39, Theorem 5, p. 181 and Example 2, p. 189] g^o can be extended for every $1 \leq p \leq \infty$ to the whole $\mathbb{R}^d$ to a function $\bar{g}^{o,p} \in \mathcal{W}_p^1(\mathbb{R}^d)$. By the equivalence between Sobolev spaces and Bessel potential spaces [39, Theorem 3, p. 135], for $1 < p < \infty$ one has $\bar{g}^{o,p} \in \mathcal{B}_p^1(\mathbb{R}^d)$, which, when restricted to X , proves the statement.

(ii) The value function J^o is continuous on X by a direct application of [40, Theorem 4.3, p. 72 and Theorem 4.6, p. 79]. As g^o is the optimal policy, we have $J^o(x) = h(x, g^o(x)) + \beta J^o(g^o(x))$. Then, by Benveniste-Scheinkman's envelope theorem [14, Corollary 4.1, p. 90] for $x \in \text{int}(X)$ we get $\nabla J^o(x) = \nabla_1 h(x, g^o(x))$. So, as $h \in \mathcal{C}^2(D)$ and g^o is Lipschitz continuous, J^o has Lipschitz continuous partial derivatives on $\text{int}(X)$. Thus, $J^o \in \mathcal{W}_\infty^2(\text{int}(X))$ and for $1 < p < \infty$ the proof proceeds as in (i). For $p = 1$, one can proceed analogously, recalling that for $d = 1$

the equivalence between $\mathcal{W}_1^2(\mathbb{R}^d)$ and $\mathcal{B}_1^2(\mathbb{R}^d)$ still holds (see the Appendix), whereas for $d > 1$ one has the inclusion $\mathcal{W}_1^2(\mathbb{R}^d) \subset \mathcal{B}_1^2(\mathbb{R}^d)$ (see the Appendix). ■

5.3. A local smoothness result. Sometimes it is not necessary to have a good approximation of J^o on the whole X . For example, for the multidimensional case, in [14, Theorem 8.2, p. 159] conditions were given for the existence of an open set $\Omega \subset \mathbb{R}^d$ such that $g^o(\Omega) \subset \Omega$; so, when $x \in \Omega$, the maximization step in ADP can be performed assuming that the next state y is in Ω . So, also local results on the smoothness of the value and optimal policy functions are of interest.

Assumption 5.3. *The transition reward $h \in \mathcal{C}^2(\text{int}(D))$.*

Assumption 5.4. *There exists an open set $\Omega \subset X$ such that*

- (i) an optimal policy g^o exists over Ω and is continuous therein;*
- (ii) the map g^o is uniformly interior, i.e., there exists $A_1 > 0$ such that $\|y - g^o(x)\| < A_1$ and $x \in \Omega \Rightarrow y \in D(x) \cap \Omega$;*
- (iii) the value function J^o is finite on Ω .*

Assumption 5.5. *For the same set Ω and the same constant A_1 as in Assumption 5.4, the following hold:*

- (i) there exists $A_2 > 0$ such that, for every $a, b \in \mathbb{R}^d$,*

$$\|y - g^o(x)\| < A_1 \text{ and } x \in \Omega \Rightarrow (a, b)' \nabla^2 h(x, y) (a, b) \leq -A_2 \|a\|^2;$$

- (ii) there exist $A_3, A_4 \geq 0$ such that, for every $a \in \mathbb{R}^d$*

$$\|y - g^o(x)\| < A_1 \text{ and } x \in \Omega \Rightarrow a' \nabla_{1,1}^2 h(x, y) a \geq -A_3 \|a\|^2 \text{ and } \|\nabla_{1,2}^2 h(x, y)\| \leq A_4.$$

Moreover, $\nabla^2 h(x, y)$ is uniformly continuous over the set $\{(x, y) : x \in \Omega \text{ and } \|y - g^o(x)\| < A_1\}$.

A bounded set $\Omega \subset \mathbb{R}^d$ has $\mathcal{C}^1$ boundary if its boundary can be $\mathcal{C}^1$ -embedded in $\mathbb{R}^d$ (see [15, p. 34] for a formal definition). For a set $\Omega \subseteq \mathbb{R}^d$, we denote by $\text{cl}(\Omega)$ its closure in the Euclidean topology.

Proposition 5.6. *Let Assumptions 5.1 and 5.3 hold, $\Omega \subset X$ be an open set that satisfies Assumptions 5.4 and 5.5. Let $\Omega_1 \subset \Omega$ be open, with $\mathcal{C}^1$ boundary, and such that $\text{cl}(\Omega_1) \subset \Omega$. Then the following hold.*

- (i) For every $1 < p < \infty$ there exists a function $\bar{g}^{o,p} \in \mathcal{B}_p^1(\mathbb{R}^d)$ such that $g^o|_{\Omega_1} = \bar{g}^{o,p}|_{\Omega_1}$.
- (ii) For every $1 \leq p < \infty$ there exists a function $\bar{J}^{o,p} \in \mathcal{B}_p^2(\mathbb{R}^d)$ such that $J^o|_{\Omega_1} = \bar{J}^{o,p}|_{\Omega_1}$.

Proof. (i) Since the hypotheses of [31, Theorem 1] are satisfied, $g^o \in \mathcal{C}^1(\Omega)$. Then for every $1 \leq p \leq \infty$, $g^o \in \mathcal{W}_p^1(\Omega_1)$. As in the proof of Proposition 5.2 (i), for $1 < p < \infty$ we conclude by applying Sobolev's extension theorem (which can be applied here thanks to the smoothness of the boundary of Ω_1 ; see [39, Example 1, p. 189]) and the equivalence between Sobolev spaces and Bessel potential spaces.

(ii) $J^o \in \mathcal{C}^2(\Omega)$ by [31, Theorem 1]. Then for $1 < p < \infty$ we proceed as in (i), and we extend the result to the case $p = 1$ as in the proof of Proposition 5.2 (ii). ■

Remark 5.7. The literature provides other local results that guarantee higher-order differentiability of the value and optimal policy functions for Problem Σ_∞ . Here we recall one of them. Let $m \geq 2$ be a positive integer. In [14, Theorem 8.2, p. 159] it was proved that if the transition reward h is concave and of class $\mathcal{C}^m(D)$ and some other technical assumptions hold, then there exists an open neighbourhood Ω of an interior fixed point³ x^* of g^o such that the optimal policy function g^o is of class $\mathcal{C}^{m-1}(\Omega)$. Thus, proceeding as in [14, p. 174], by Benveniste-Scheinkman's envelope theorem and the chain rule one can conclude that the (concave) value function is of class $\mathcal{C}^m(\Omega)$. Then, applying the same proof technique used in Proposition 5.6, one can prove that for a suitable open set $\Omega_1 \subset \Omega$ and for $1 < p < \infty$ there exists $\bar{J}^{o,p} \in \mathcal{B}_p^m(\mathbb{R}^d)$ such that $J^o|_{\Omega_1} = \bar{J}^{o,p}|_{\Omega_1}$. If m is even, this holds also for $p = 1$ (see the Appendix). A similar statement holds for the optimal policy g^o .

5.4. Comparison with smoothness results for the finite-horizon case. It is interesting to compare the smoothness results derived for Problem Σ_∞ with those for its following finite-horizon version, stated for a positive integer N .

³The existence of such an interior fixed point x^* can be checked using [14, Proposition 7.3, p. 133].

$$\begin{aligned} \text{Problem } \Sigma_N : \quad J^o(x_0) &= \sup \left\{ \sum_{t=0}^{N-1} \beta^t h_t(x_t, x_{t+1}) + \beta^N h_N(x_N) \right\} \\ \text{s.t.} \quad x_t &\in X_t, (x_t, x_{t+1}) \in D_t, \quad t = 0, 1, \dots, N-1. \end{aligned}$$

The function $h_N : X_N \rightarrow \mathbb{R}$ is a *final reward*, which depends only on the final state x_N . As done for Problem Σ_∞ , for simplicity of notations and without loss of generality, in the following we assume that d_t , X_t , h_t and D_t do not depend on the time t , so we shall use the notations d , X , h and D . Let us consider the following subproblems, where the function $J_t^o : X \rightarrow \mathbb{R}$ is called *t-th value function*:

$$\begin{aligned} (5) \quad J_N^o(x_N) &= h(x_N) \\ J_t^o(x_t) &= \sup \left\{ \sum_{k=t}^{N-1} \beta^{k-t} h(x_k, x_{k+1}) + \beta^{N-t} h_N(x_N) \right\} \\ \text{s.t.} \quad x_t &\in X, x_k, x_{k+1} \in D, \quad k = t, \dots, N-1 \\ t &= N-1, \dots, 0. \end{aligned}$$

The subproblems (5) can be solved recursively by using Bellman's operator. Indeed,

$$J_t^o(x_t) = (TJ_{t+1}^o)(x_t), \quad t = N-1, \dots, 0.$$

Similarly, the optimal policy functions are given by

$$g_t^o(x) \in \operatorname{argmax}_{y \in D(x)} [h(x, y) + \beta J_{t+1}^o(y)], \quad t = N-1, \dots, 0.$$

For Problem Σ_N , global smoothness results for the t -th value and optimal policy functions are available (see next Proposition 5.10), which are stronger than those derived here for the value and optimal policy functions in Problem Σ_∞ . Indeed, Proposition 5.2 gives a global smoothness result, which however is limited to Lipschitz continuity of the optimal policy function and Lipschitz continuity of the partial derivatives of the value function. On the other hand, the smoothness results for Problem Σ_∞ that can be obtained following Remark 5.7 have a larger degree of smoothness than Proposition 5.2, but are of local nature.

We shall need the following technical lemma. Given a partitioned real matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

such that D is non-singular, *Schur's complement* M/D of D in M is defined [44, p. 18] as the matrix

$$M/D = A - BD^{-1}C.$$

For a symmetric real matrix, we denote by $\lambda_{\max}$ its maximum eigenvalue.

Lemma 5.8. *Let $M = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix}$ be a partitioned symmetric negative-semidefinite matrix such that D is non-singular. Then $\lambda_{\max}(M/D) \leq \lambda_{\max}(M)$.*

Proof. It readily follows by [44, Theorem 2.13, p. 69] and the example given in [44, p. 70]. ■

The next proposition gives conditions under which the t -th value and optimal policy functions of Problem Σ_N can be extended on the whole $\mathbb{R}^d$ to functions in the Bessel potential spaces $\mathcal{B}_p^m(\mathbb{R}^d)$ and $\mathcal{B}_p^{m-1}(\mathbb{R}^d)$, respectively.

For $X \subseteq \mathbb{R}^d$ convex and $\alpha \in \mathbb{R}$, a function $f : X \rightarrow \mathbb{R}$ is α -concave on X if $f(x) + \frac{1}{2}\alpha\|x\|^2$ is concave on X [30]. If f is of class $\mathcal{C}^2(X)$, then a sufficient condition for its α -concavity is

$$\sup_{x \in X} \lambda_{\max}(\nabla^2 f(x)) \leq -\alpha,$$

where $\lambda_{\max}(\nabla^2 f(x))$ is the maximum eigenvalue of the Hessian $\nabla^2 f(x)$.

Assumption 5.9. *Let $m \geq 2$ be integer. Then the following hold:*

- (i) *there exist optimal policies $g_{N-1}^o, \dots, g_0^o$ that are continuous and interior on X , i.e., for every $x_t \in X$, $g_t^o(x_t) \in \text{int}(D(x_t))$;*
- (ii) *$h_N \in \mathcal{C}^m(X)$ is concave, $h \in \mathcal{C}^m(D)$, and there exists $\alpha > 0$ such that h is α -concave.*

Proposition 5.10. *Let Assumptions 5.1 and 5.9 hold. Then for every integer $t = 0, \dots, N$ and every $1 < p < \infty$ the following hold.*

- (i) *$J_t^o \in \mathcal{C}^m(X)$ and there exists $\bar{J}_t^{o,p} \in \mathcal{B}_p^m(\mathbb{R}^d)$ such that $J_t^o = \bar{J}_t^{o,p}|_X$; when m is even, it holds also for $p = 1$.*

(ii) $g_t^o \in \mathcal{C}^{m-1}(X)$ and there exists $\bar{g}_t^{o,p} \in \mathcal{B}_p^{m-1}(\mathbb{R}^d)$ such that $g_t^o = \bar{g}_t^{o,p}|_X$; when $m \geq 3$ is odd, it holds also for $p = 1$.

Proof. (i) Taking the hint from the proof of [31, Theorem 1], let us first show by backward induction that $J_t^o \in \mathcal{C}^m(X)$ and $g_t^o \in \mathcal{C}^{m-1}(X)$, $t = 0, \dots, N$. Since $J_N^o = h_N$, one has that $J_N^o \in \mathcal{C}^m(X)$ and is concave by hypothesis. Now, fix t and suppose that $J_{t+1}^o \in \mathcal{C}^m(X)$ and is concave. Since by hypothesis the optimal policy g_t^o is interior, the first-order optimality condition $\nabla_2 h(x_t, g_t^o(x_t)) + \beta \nabla J_{t+1}^o(g_t^o(x_t)) = 0$ holds. By the implicit function theorem we get

$$(6) \quad \nabla g_t^o(x_t) = -[\nabla_{2,2}^2(h(x_t, g_t^o(x_t))) + \beta \nabla^2 J_{t+1}^o(g_t^o(x_t))]^{-1} \nabla_{2,1}^2 h(x_t, g_t^o(x_t)),$$

where the matrix $\nabla_{2,2}^2(h(x_t, g_t^o(x_t))) + \beta \nabla^2 J_{t+1}^o(g_t^o(x_t))$ is non-singular by the α -concavity of h , the concavity of J_{t+1}^o , and their twice differentiability. By differentiating the two members of (6) up to derivatives of h of order m , it follows⁴ that $g_t^o \in \mathcal{C}^{m-1}(X) \subset \mathcal{W}_p^{m-1}(\text{int}(X))$ for every $p \geq 1$.

By differentiating the equality $J_t^o(x_t) = h(x_t, g_t^o(x_t)) + \beta J_{t+1}^o(g_t^o(x_t))$, we obtain

$$\nabla J_t^o(x_t) = \nabla_1 h(x_t, g_t^o(x_t)) + [\nabla_2 h(x_t, g_t^o(x_t)) + \beta \nabla J_{t+1}^o(g_t^o(x_t))]' \nabla g_t^o(x_t).$$

So by the first-order optimality condition we get

$$(7) \quad \nabla J_t^o(x_t) = \nabla_1 h(x_t, g_t^o(x_t)).$$

By differentiating the two members of (7) up to derivatives of h of order m , it follows that $J_t^o \in \mathcal{C}^m(X) \subset \mathcal{W}_p^m(\text{int}(X))$ for every $p \geq 1$.

In order to conclude the backward induction step, it remains to show that J_t^o is concave. By differentiating (7) and using (6), the Hessian of J_t^o has the expression

$$\begin{aligned} \nabla^2 J_t^o(x_t) &= \nabla_{2,2}^2 h(x_t, g_t^o(x_t)) + \\ &\quad - \nabla_{1,2}^2 h(x_t, g_t^o(x_t)) [\nabla_{2,2}^2 h(x_t, g_t^o(x_t)) + \beta \nabla^2 J_{t+1}^o(x_t, g_t^o(x_t))]^{-1} \nabla_{2,1}^2 h(x_t, g_t^o(x_t)), \end{aligned}$$

⁴One can show that all the partial derivatives up to the order $m - 1$ of the components of g^o are bounded on X . Indeed, if $A(y)$ is a symmetric nonsingular matrix dependent on a vector parameter y , an expression for the partial derivatives of its inverse $A^{-1}(y)$ can be obtained by differentiating the identity $A(y)A^{-1}(y) = I$. Then the absolute values of the partial derivatives of $A^{-1}(y)$ can be uniformly bounded from above via uniform bounds on the eigenvalues of $A^{-1}(y)$.

which is the Schur's complement of $[\nabla_{2,2}^2 h(x_t, g_t^o(x_t)) + \beta \nabla^2 J_{t+1}^o(x_t, g_t^o(x_t))]$ in the matrix

$$\begin{pmatrix} \nabla_{2,2}^2 h(x_t, g_t^o(x_t)) & \nabla_{1,2}^2 h(x_t, g_t^o(x_t)) \\ \nabla_{2,1}^2 h(x_t, g_t^o(x_t)) & \nabla_{2,2}^2 h(x_t, g_t^o(x_t)) + \beta \nabla^2 J_{t+1}^o(x_t, g_t^o(x_t)) \end{pmatrix}.$$

Such a matrix is negative-semidefinite by the concavity and twice differentiability of h and J_{t+1}^o . Then, by Lemma 5.4 J_t^o is concave.

The remaining of the proof proceeds as the proof of Proposition 5.2 (i) for $1 < p < \infty$ and the proof of Proposition 5.2 (ii) for $p = 1$, recalling that for $d = 1$ the equivalence between $\mathcal{W}_1^m(\mathbb{R}^d)$ and $\mathcal{B}_1^m(\mathbb{R}^d)$ still holds for m even (see the Appendix), whereas for $d > 1$ and m even, one has the inclusion $\mathcal{W}_1^m(\mathbb{R}^d) \subset \mathcal{B}_1^m(\mathbb{R}^d)$ (see the Appendix).

(ii) As proved in (i), $g_t^o \in \mathcal{C}^{m-1}(X) \subset \mathcal{W}_p^{m-1}(\text{int}(X))$ for every $p \geq 1$. Then one proceeds as at the end of the proof of (i). ■

Remark 5.11. Inspection of the proof of Proposition 5.10 shows that its arguments allow to obtain an a-priori (i.e., independent on the specific form of J_{t+1}^o and g_t^o) bounds on the eigenvalues of $[\nabla_{2,2}^2(h(x_t, g_t^o(x_t))) + \beta \nabla^2 J_{t+1}^o(g_t^o(x_t))]$, hence on the absolute values of the partial derivatives (up to the order $m - 1$) of the components of g_t^o . A-priori upper bounds on the absolute values of the partial derivatives of J_t^o up to the order m can be obtained, too.

Remark 5.12. A result very similar to Proposition 5.10 is given in [21, 23]. However, its proof is slightly different since it does not use Schur's complement.

6. VALUE OR POLICY FUNCTION APPROXIMATION?

On the basis of the analysis of the previous sections, the following pros and cons of the two approaches can be emphasized.

6.1. Pros of value-function approximation.

- It is based (although in an approximate form) on Bellman's principle of optimality, which a priori excludes some solutions from the search.
- Usually, the degree of smoothness of the value function is higher than the degree of smoothness of the optimal policy. This is in favor of value-function

approximation, since typically a higher smoothness makes the approximation task easier.

6.2. Pros of policy-function approximation.

- Optimal policy function approximation can be applied to general functional optimization problems [28, 45, 20], independently on the possibility of dividing the decisional process into stages. So, it is a more general methodology than ADP.
- The main advantage is that it does not require to maximize an approximation of the value function; this, instead, has to be done in ADP. Unless suitable smoothness properties are known a priori, in value-function approximation this maximization step requires an approximation in the supremum norm (otherwise, one loses the control on the error propagation from one iteration to the next one). In policy function approximation, instead, in principle the error analysis can be done also in terms of $\mathcal{L}_p$ -norms with $p \neq \infty$. Measuring the approximation error in $\mathcal{L}_p$ -norms is important, e.g., for the following reasons. (i) In the literature about nonlinear approximation theory, several upper bounds on the worst-case approximation error have been derived in $\mathcal{L}_p$ norms with $p \neq \infty$ (see [4, 13, 16, 19, 26, 27, 29, 33] and the references therein). (ii) When approximating families containing parameters to be optimized are used (e.g., feedforward neural networks), typically the values of such parameters are determined by nonlinear programming algorithms (e.g., [1, 3, 7, 24, 43]). The use, e.g., of the $\mathcal{L}_2$ norm instead of the supremum norm, allows the application of gradient-descent methods.

The following is a straightforward example of an approximation bound in $\mathcal{L}_1$ -norm. For Problem Σ_N with one stage (i.e., $N = 1$), Lipschitz continuous transition reward h , and Lipschitz continuous final reward h_1 , one

has

$$\begin{aligned}
& \|J^o - \tilde{J}^o\|_{\mathcal{L}_1(X)} \leq \int_X (\|h(x_0, g^o(x_0)) - h(x_0, \tilde{g}^o(x_0))\| \\
& + \|h_1(g^o(x_0)) - h_1(\tilde{g}^o(x_0))\|) dx_0 \\
& \leq \int_X (L_h \|g^o(x_0) - \tilde{g}^o(x_0)\| + L_{h_1} \|g^o(x_0) - \tilde{g}^o(x_0)\|) dx_0 \\
& = (L_h + L_{h_1}) \|g^o - \tilde{g}^o\|_{\mathcal{L}_1(X)}.
\end{aligned}$$

Then, an approximation upper bound on $\|g^o - \tilde{g}^o\|_{\mathcal{L}_1(X)}$ implies an approximation upper bound on $\|J^o - \tilde{J}^o\|_{\mathcal{L}_1(X)}$. However, the extension of this example to multistage problems is not immediate.

- For Problem Σ_∞ and $\frac{1}{2} \leq \beta < 1$, the upper bound given by Proposition 3.3 suffers from sort of “curse of time horizon” [21], i.e. it grows exponentially fast with N , unless $0 < \beta < \frac{1}{2}$. This drawback can be mitigated by resorting to the “multistage lookahead” technique [10, Section 6.1.2], at the cost of an additional computational burden per iteration [21].

7. EXTENSIONS

Let us consider an alternative formulation of Problem Σ_∞ , typically adopted in optimal control theory. In that field, the transition rewards are usually expressed in terms of the current state x_t and the currently applied *control* u_t , i.e., in the form $h(x_t, u_t)$. Accordingly, instead of the correspondence D , one considers a *dynamical system* modeled by a *state equation* $x_{t+1} = f(x_t, u_t)$, where $f : X \times U \rightarrow \mathbb{R}^d$ is a *state transition function* (i.e., the *system dynamics*) and $U \subseteq \mathbb{R}^s$ is the set to which the control vector belongs (*control space*).

As discussed in [14, Section 5.4.3, pp. 101-103], in the deterministic case the two formulations (i.e., the one in terms of correspondences and the one in terms of dynamical systems) are equivalent (in the sense that it is always possible to move from one to the other, by choosing a suitable dynamical system and a suitable correspondence) but usually this equivalence does not preserve concavity of the transition rewards. Indeed, if h is concave when it is expressed as a function of x_t and x_{t+1} , it may not be concave when it is expressed as a function of x_t and u_t in the dynamical system formulation. The same considerations apply when one moves from the second formulation to the first one.

However, a remarkable exception occurs for linear dynamical systems with convex constraints on the pair (x_t, u_t) : concavity of h is preserved in both senses, as it can be easily proved by using the results of [12, Section 3.2.2] and [9, Proposition 2.3.6], respectively. So, in such a case our results can be applied also to dynamic optimization problems formulated in terms of dynamical systems. Another possible extension consists in restating our results directly in terms of the latter formulation (i.e., without using the formulation in terms of correspondences as an intermediate step).

APPENDIX

The following relationships between Sobolev spaces and Bessel potential spaces hold in the cases $p = 1$ and $p = \infty$ and are stated without proof in [39, Remark 6.6, p. 160]. For the case $1 < p < \infty$, see [39, Theorem 3, p. 135].

Theorem 7.1. (i) Let $d = 1$, m even, and $p = 1$ or $p = \infty$; then $\mathcal{W}_p^m(\mathbb{R}) = \mathcal{B}_p^m(\mathbb{R})$.
(ii) Let $d > 1$, m even, and $p = 1$ or $p = \infty$; then $\mathcal{W}_p^m(\mathbb{R}^d) \subseteq \mathcal{B}_p^m(\mathbb{R}^d)$ (the inclusion is strict for both $p = 1$ and $p = \infty$).

For completeness, in the following we give a sketch of the proof of the inclusion $\mathcal{W}_p^m(\mathbb{R}^d) \subseteq \mathcal{B}_p^m(\mathbb{R}^d)$ in Theorem 7.1 (ii). By $\mathcal{C}_0^\infty(\mathbb{R}^d)$ we denote the space of functions on $\mathbb{R}^d$ that are infinitely times differentiable and have compact support.

Sketch of proof. (a) Let $d = 1$ and $m = 2$. We show that $\mathcal{W}_1^2(\mathbb{R}) \subseteq \mathcal{B}_1^2(\mathbb{R})$. Let $u \in \mathcal{W}_1^2(\mathbb{R})$; then, u and its weak derivatives $\frac{du}{dx}$ and $\frac{d^2u}{dx^2}$ are in $\mathcal{L}_1(\mathbb{R})$. Being functions in $\mathcal{L}_1(\mathbb{R})$ they admit a Fourier transform. As u and $\frac{du}{dx}$ are absolutely continuous [15, Lemma 1.38, p. 32], one has $\widehat{\left(\frac{d^2u}{dx^2}\right)}(s) = -|s|^2\hat{u}(s)$. Consider the function $q \in \mathcal{B}_1^2(\mathbb{R})$ such that $q = \sqrt{2\pi} \left(u - \frac{d^2u}{dx^2}\right) * G_2$. Since $\left(u - \frac{d^2u}{dx^2}\right) \in \mathcal{L}_1(\mathbb{R})$, $G_2 \in \mathcal{L}_1(\mathbb{R})$ by [39, Proposition 2, p. 132], and $q \in \mathcal{L}_1(\mathbb{R})$ by [36, Theorem 8.14, p. 170], then by [36, Theorem 9.2 (c), p. 179] one has $\hat{q}(s) = \sqrt{2\pi} \widehat{\left(u - \frac{d^2u}{dx^2}\right)}(s) \hat{G}_2(s) = \hat{u}(s)$. As u and q are in $\mathcal{L}_1(\mathbb{R})$ and have the same Fourier transform, by [36, Theorem 9.12, p. 185] they coincide as elements of $\mathcal{L}_1(\mathbb{R})$. Note that this gives also the following relationship between Bessel norm and Sobolev norm: $\|u\|_{\mathcal{B}_1^2(\mathbb{R})} = \sqrt{2\pi} \left\| u - \frac{d^2u}{dx^2} \right\|_{\mathcal{L}_1(\mathbb{R})}$, where the second term can be bounded from above in terms of one of the equivalent norms of $\mathcal{W}_1^2(\mathbb{R})$.

(b) Let $d > 1$ and $m = 2$. We show that $\mathcal{W}_1^2(\mathbb{R}^d) \subseteq \mathcal{B}_1^2(\mathbb{R}^d)$. In the following we use the same notation for classical and weak partial derivatives. As $\mathcal{C}_0^\infty(\mathbb{R}^d)$ is dense in $\mathcal{W}_1^2(\mathbb{R}^d)$ [2, Theorem 3.18, p. 54 and Corollary 3.19, p. 56], by Riemann-Lebesgue's lemma stated for the d -dimensional case [38, Theorem 0.1.2, p. 2] the well-known relationship used in (a) among the Fourier transforms of a function in $\mathcal{C}_0^\infty(\mathbb{R}^d)$ and of its partial derivatives extends to the multidimensional case [38, Theorem 0.14, p. 2]. Then, for $u \in \mathcal{C}_0^\infty(\mathbb{R}^d)$, one obtains $(2\pi)^{\frac{d}{2}} \left(u - \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2} \right) * G_2 = u$. As $\mathcal{C}_0^\infty(\mathbb{R}^d)$ is dense in $\mathcal{W}_1^2(\mathbb{R}^d)$ and, for d -dimensional convolutions, $q = f * g$ implies $\|q\|_{\mathcal{L}_1(\mathbb{R}^d)} \leq \|f\|_{\mathcal{L}_1(\mathbb{R}^d)} \|g\|_{\mathcal{L}_1(\mathbb{R}^d)}$ [2, Theorem 4.30, p. 90], we get $(2\pi)^{\frac{d}{2}} \left(u - \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2} \right) * G_2 = u$ for all $u \in \mathcal{W}_1^2(\mathbb{R}^d)$. Finally one has $\|u\|_{\mathcal{B}_1^2(\mathbb{R}^d)} = (2\pi)^{\frac{d}{2}} \left\| u - \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2} \right\|_{\mathcal{L}_1(\mathbb{R}^d)}$ for all $u \in \mathcal{W}_1^2(\mathbb{R}^d)$, where the second term can be bounded from above in terms of one of the equivalent norms of $\mathcal{W}_1^2(\mathbb{R}^d)$.

(c) Similarly, in the general case $d \geq 1$ and m even, one gets $\mathcal{W}_1^m(\mathbb{R}^d) \subseteq \mathcal{B}_1^m(\mathbb{R}^d)$ and $\|u\|_{\mathcal{B}_1^m(\mathbb{R}^d)} = \|v\|_{\mathcal{L}_1(\mathbb{R}^d)}$ for all $u \in \mathcal{W}_1^m(\mathbb{R}^d)$, where v is the inverse Fourier transform of $\hat{v} = (2\pi)^{\frac{d}{2}} (1 + \|s\|^2)^{\frac{m}{2}} \hat{u}$ (note that, as m is even, v can be expressed as a weighted sum of the partial derivatives of u up to the order m).

REFERENCES

- [1] E. Aarts and J. Korst, *Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*, John Wiley & Sons, 1989.
- [2] R. A. Adams, *Sobolev spaces*, Academic Press, NY, 1975.
- [3] A. Alessandri, M. Sanguineti, and M. Maggiore, *Optimization-based learning with bounded error for feedforward neural networks*, IEEE Trans. on Neural Networks **13** (2002), 261–273.
- [4] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. on Information Theory **39** (1993), 930–945.
- [5] R. Bellman, *Dynamic programming*, Princeton University Press, Princeton, NJ, 1957.
- [6] L. M. Benveniste and J. A. Scheinkman, *On the differentiability of the value function in dynamic models of economics*, Econometrica **47** (1979), 727–732.
- [7] D. P. Bertsekas, *A new class of incremental gradient methods for least squares problems*, SIAM J. on Optimization **7** (1997), 913–926.
- [8] D. P. Bertsekas, *Dynamic programming and optimal control, vol. 1*, Athena Scientific, Belmont, MA, 2005.
- [9] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Belmont, MA, 2003.
- [10] D. P. Bertsekas and J. Tsitsiklis, *Neuro-dynamic programming*, Athena Scientific, Belmont, MA, 1996.

- [11] J. Blot and B. Crettez, *On the smoothness of optimal paths*, Decisions in Economics and Finance **27** (2004), 1–34.
- [12] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [13] L. Breiman, *Hinging hyperplanes for regression, classification and function approximation*, IEEE Trans. on Information Theory **39** (1993), 999–1013.
- [14] F. Cugno and L. Montrucchio, *Scelte intertemporali. Teoria e modelli*, Carocci Editore, Roma, 1998.
- [15] B. Dacorogna, *Introduction to the calculus of variations*, Imperial College Press, 2004.
- [16] C. Darken, M. Donahue, L. Gurvits, and E. Sontag, *Rate of approximation results motivated by robust neural network learning*, Proc. of the Sixth Annual ACM Conference on Computational Learning Theory, pp. 303–309. The Association for Computing Machinery, New York, 1993.
- [17] I. Ekeland and T. Turnbull, *Infinite-dimensional optimization and convexity*, The University of Chicago Press, Chicago, USA, 1983.
- [18] D. Gale, *On optimal development in a multi-sector economy*, Review of Economic Studies **34** (1967), 1–18.
- [19] F. Girosi and G. Anzellotti, *Rates of convergence for Radial Basis Functions and neural networks*. In R. J. Mammone, Ed., Artificial Neural Networks for Speech and Vision, pp. 97–113. Chapman & Hall, 1993.
- [20] S. Giulini and M. Sanguineti, *Approximation schemes for functional optimization problems*, J. of Optimization Theory and Applications, DOI:10.1007/s10957-008-9471-6 (2008).
- [21] G. Gnecco and M. Sanguineti, *Accuracy of suboptimal solutions to dynamic optimization problems by approximate dynamic programming*, DIST Internal Report 2008, submitted.
- [22] G. Gnecco and M. Sanguineti, *Approximation error bounds via Rademacher complexity*, Applied Mathematical Sciences, vol. 2, pp. 153–176, 2008.
- [23] G. Gnecco and M. Sanguineti, *Suboptimal solutions to dynamic optimization problems via approximations of the policy functions*, J. of Optimization Theory and Applications, to appear.
- [24] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, 1989.
- [25] J. Heinonen, *Lectures on analysis on metric spaces*, Springer, New York, 2001.
- [26] V. Kůrková and M. Sanguineti, *Bounds on rates of variable-basis and neural-network approximation*, IEEE Trans. on Information Theory **47** (2001), 2659–2665.
- [27] V. Kůrková and M. Sanguineti, *Comparison of worst-case errors in linear and neural network approximation*, IEEE Trans. on Information Theory **48** (2002), 264–275.
- [28] V. Kůrková and M. Sanguineti, *Error estimates for approximate optimization by the extended Ritz method*, SIAM J. on Optimization **18** (2005), 461–487.
- [29] V. Kůrková and M. Sanguineti, *Geometric upper bounds on rates of variable-basis approximation*, IEEE Trans. on Information Theory, **54**, December 2008 (DOI: 10.1109/TIT.2008.2006383).

- [30] L. Montrucchio, *Lipschitz continuous policy functions for strongly concave optimization problems*, J. of Mathematical Economics **16** (1987), 259–273.
- [31] L. Montrucchio, *Thompson metric, contraction property and differentiability of policy functions*, J. of Economic Behavior and Organization **33** (1998), 449–466.
- [32] L. Montrucchio and M. Boldrin, *On the indeterminacy of capital accumulation paths*, J. of Economic Behavior and Organization **40** (1986), 26–36.
- [33] A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta Numerica **8** (1999), 143–195.
- [34] W. B. Powell, *Approximate dynamic programming. Solving the curses of dimensionality*, John Wiley & Sons, Inc., Hoboken, NJ, 2007.
- [35] M. L. Puterman, *Markov decision processes*, Wiley, 1994.
- [36] W. Rudin, *Real and complex analysis*, McGraw-Hill, 1970.
- [37] J. E. Smith and K. F. McCardle, *Structural properties of stochastic dynamic programs*, Operations Research **50** (2002), 794–809.
- [38] D. Sogge, *Fourier integrals in classical analysis*, Cambridge University Press, 1993.
- [39] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ, 1970.
- [40] N. L. Stokey, R. E. Lucas, and E. Prescott, *Recursive methods in economic dynamics*, Harvard University Press, MA, 1989.
- [41] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [42] A. Venditti, *Nonlinear dynamics and indeterminacy in multisector growth models*, XIIIth Summer School of the European Economic Association on "Endogenous Fluctuations", Portuguese Catholic University, Lisbon, Portugal, 2002.
- [43] G. Yin, *Rates of convergence for a class of global stochastic optimization algorithms*, SIAM J. on Optimization **10** (1999), 99–120.
- [44] F. Zhang (ed.), *The Schur complement and its applications*, Springer, 2005.
- [45] R. Zoppoli, M. Sanguineti, and T. Parisini, *Approximating networks and extended Ritz method for the solution of functional optimization problems*, J. of Optimization Theory and Applications **112** (2002), 403–439.