No Curse of Dimensionality for Contraction Fixed Points Even in the Worst Case

H. Woźniakowski, *Computer Science, Columbia University and Institute of Applied Mathematics and Mechanics, University of Warsaw*[*]
J. F. Traub, *Computer Science, Columbia University*
J. Rust, *Economics, Yale University*

December 4, 1998

Abstract

We consider the problem of computing approximations to fixed points of *quasilinear contraction mappings* Γ defined on the space of continuous functions of d variables. Our main emphasis is on large d. Examples of such mappings include the *Bellman operator* from the theory of dynamic programming. This paper proves that there exist *deterministic algorithms* for computing approximations to fixed points for some classes of quasilinear contraction mappings which are *strongly tractable,* i.e., in the worst case the number of function evaluations $n(\epsilon, d)$ needed to compute an ε-approximation to the solution of $V = \Gamma(V)$ at any finite number of points in its domain is bounded by $C\varepsilon^{-p}$ where both C and p are independent of d. This is done by using relations between the quasilinear contraction problem and the conditional expectation and approximation problems. The conditional expectation problem is equivalent to weighted multivariate integration. This allows us to apply recent proof technique and results on the strong tractability of weighted multivariate integration and approximation to establish strong tractability for the quasilinear fixed point problem. In particular, this holds when the fixed points belong to a Sobolev space for a specific weighted norm.

---

[*] This research was supported in part by the Alfred Sloan Foundation.

1 Introduction

This paper considers the problem of computing approximations to fixed points for a class of *quasilinear contraction mappings*. The contraction mappings are defined as certain operators of linear conditional expectations. Fixed point problems in this class arise frequently in engineering and economic applications including dynamic programming ([3], [4], [7], [24]), econometrics ([23]), and asset pricing ([13], [22]). The goal is to compute an approximation to the unique fixed point $V = \Gamma(V)$ of the quasilinear contraction mapping Γ, where V is an element of the space B_d of continuous functions of d variables defined on a compact set $S_d \subset \mathbb{R}^d$. We stress that the values of d considered in this paper can be arbitrarily large, and our focus is on finding sufficient conditions for which the worst case complexity of computing an ϵ-approximation to V is either independent of d, or is at most polynomial in d.

At least as far back as Bellman 1957, it has been thought that fixed point problems of the type we are considering are subject to an unavoidable "curse of dimensionality" in the worst case deterministic setting. That is, the cost of computing an ϵ-approximation to V, $\mathrm{cost}(\varepsilon, d)$, increases exponentially fast in d in the worst case. In computer science, problems that are subject to the curse of dimensionality are called *intractable* whereas problems where $\mathrm{cost}(\varepsilon, d)$ can be bounded by a polynomial in ε^{-1} and d are said to be *tractable.* When $\mathrm{cost}(\varepsilon, d)$ can be bounded by a polynomial in ε^{-1} independently of d then they are called *strongly tractable* and the minimal exponent of ε^{-1} is called the *strong exponent*, see [35] for a precise definition.

Previous analyses of the complexity of the contraction fixed point problem (e.g. [27]) considered contraction mappings on finite-dimensional spaces (e.g. $\mathbb{R}^m$) or assumed that $\Gamma(W)$ could be exactly computed at unit cost. In this paper we study the contraction fixed point problem on infinite-dimensional spaces for which $\Gamma(W)$ can only be approximately computed. We consider the cost of approximating $\Gamma(W)$ as part of the overall cost of the fixed point algorithm. We consider a class of contraction mappings for which Γ is defined as a particular nonlinear functional of a finite number of (linear) conditional expectation operators, E_k, $k = 1, \ldots, m$ — hence the name "quasilinear". To evaluate $\Gamma(W)(s)$ for an arbitrary $W \in B_d$ and $s \in S_d$ we need to compute the functions $E_k(W)$ which ostensibly involve computing a continuum of weighted multivariate integrals of d variables. Since d is assumed to be large, even the problem of approximating a single weighted multivariate integral can be challenging.

We study successive approximation algorithms for approximating V of the form $V_{i+1} = \hat{\Gamma}_i(V_i)$, where $\hat{\Gamma}_i$ is an approximate contraction mapping constructed from approximate conditional expectation operators $\hat{E}_{k,i}$ using quadrature rules that approximate $E_{k,i}(W)$ by weighted sums of W evaluated at a finite number of points n_i in S_d, where n_i may depend on the current iteration i. Let $n(\epsilon, d)$ denote the total number of points used in the process of computing an ϵ-approximation to $V = \Gamma(V)$. We prove that there exist algorithms where $n(\epsilon, d)$ is independent of d, so in this sense the quasilinear contraction problem is strongly tractable. However even though $n(\epsilon, d)$ is independent of d, the total cost of successive approximations in terms of cpu time is generally not independent of d since the cpu time needed to evaluate an arbitrary function $W(s)$ for $s \in S_d$ generally increases in d. If this latter cost increases polynomially in d, then strong tractability of the quasilinear contraction problem measured in terms of function evaluations, $n(\epsilon, d)$, implies tractability when complexity is measured in terms of total cpu time, $\mathrm{cost}(\varepsilon, d)$.

Chow and Tsitsiklis, [5] were the first to prove that the problem of computing contraction fixed points of the type we consider is intractable. They actually considered the problem of computing approximations to the value functions V for infinite horizon Markovian decision problems (MDPs). Since the solutions to these problems are mathematically equivalent to computing the fixed point of Bellman operators, their results imply that the quasilinear contraction problem (which nests the Bellman problem as a special case), must be intractable as well. The results in [5] do not contradict our results because their worst case analysis was over a compact subset $\mathcal{F}_d$ of Lipschitz continuous functions whereas our worst case analysis is derived for a smaller set of functions, namely, the compact subset of functions $\mathcal{F}_{d,\gamma}$ of a Sobolev space equipped with a particular weighted norm, $\|f\|_{d,\gamma}$.

As discussed in [31], there are at least two basic strategies to try to "break" the curse of dimensionality:

(a) we can consider random algorithms that offer a weaker assurance on the errors,

(b) we can restrict our worst case analysis to a subset of problem instances that has some sort of "special structure" which can be exploited in order to make the problem tractable.

Both of these strategies allow us to break the curse of dimensionality of the quasilinear fixed point problem. Strategy (a) was used in [24] by Rust who showed that a randomized algorithm succeeds in breaking the curse of dimensionality for a subclass of DP problems known as *discrete decision processes* (i.e., DP problems with finite action spaces). Rust's algorithm was motivated by previous work by Tauchen and Hussey [28] on numerical methods for solving nonlinear asset pricing problems, and their work in turn was motivated by the classical Nystrom method for solving Fredholm integral equations (see, e.g., Section 18.1 of [21]). Rust's algorithm is a modified version of the Tauchen/Hussey/Nystrom algorithm by using the classical Monte Carlo algorithm for generating sample points for integration. It has long been known that multivariate integration is tractable in the randomized setting where we allow randomized algorithms, such as Monte Carlo, and we weaken the error assurance by requiring that the *expected* error is less than ε, see, e.g. [30, 31]. Rust's result can be viewed as showing that the classical Monte Carlo integration algorithm breaks the curse of dimensionality for contraction fixed point problems which have embedded multivariate integration problems that are the source of the underlying curse of dimensionality when only deterministic algorithms are allowed.

Our paper is about strategy (b). Our approach is motivated by recent work on computational complexity of multivariate integration and approximation, see [26, 33, 35, 36]. We briefly indicate the background of this work. It has long been known that multivariate integration is intractable in the worst case deterministic setting for classes of functions with a fixed degree of smoothness. However, recent computational experiments have shown that certain deterministic integration algorithms can significantly outperform the classical Monte Carlo algorithm even for very high dimensional integration, where d is as high as 360, see [20], [18], [19]. These deterministic algorithms are referred to as quasi-Monte Carlo algorithms since the only difference between them and the classical Monte Carlo algorithm is to replace the random sample points by deterministic "low discrepancy" points. Example include Sobol or generalized Faure points, see, e.g., [15], and [29]. The success of these experiments suggests that there may be subclasses of integrands that are "well behaved" or display some sort of "special structure" even for huge d that deterministic algorithms might be exploiting. If we believe that economic or engineering problems have such structure, we can hope to use such deterministic algorithms without fear of encountering the curse of dimensionality even in the worst case setting.

Sloan and Woźniakowski [26] were the first to identify a type of special structure of functions and quantify how much it can help. They showed that there exist quasi-Monte Carlo algorithms for which the curse of dimensionality is *not* present in the worst case for a Sobolev class $F_{d,\gamma}$ of functions equipped with a particular weighted norm, $\|f\|_{d,\gamma}$. The symbol γ refers to a sequence of weights $\{\gamma_i\}$ where γ_i moderates the behavior of the functions with respect to the ith variable. The γ_i enters the norm inversely so that if an element of $f \in \mathcal{F}_d$ has weighted norm, say, at most 1 then small γ_i means that the function f is almost "flat" with respect to the ith variable. If we reorder the arguments of $f(x_1, \ldots, x_d)$ to have declining weights γ_i, then this is equivalent to assuming that arguments with successively higher indices have monotonically declining effects on the values of f. If these weights converge to 0 then f effectively depends only on a few of the first variables. For example if $\gamma_i = 0$ if $i \geq k$, then f is only a function of its first k coordinates.

It was shown in [26] that the curse of dimensionality is not present for quasi-Monte Carlo algorithms iff $\sum_{i=1}^{\infty} \gamma_i < \infty$. If so, we have *strong tractability* of integration since the number $n(\varepsilon, d)$ of sample points needed to compute an ε-approximation to the true integral in the worst case setting satisfies

$$n(\varepsilon, d) \ \leq \ C \, \varepsilon^{-p} \tag{1}$$

where both C and p are constants independent of d and $p \in [1, 2]$. This result shows that for problems with this type of special structure, the curse of dimensionality is not present even in the worst case setting. Previous "conventional wisdom" was that only variants of the classical Monte Carlo algorithm can be effective for high dimensional integration problems, see, e.g. [6].

The proof in [26] is nonconstructive, i.e., it is an existence result that doesn't specify for which quasi-Monte Carlo algorithm the bound (1) is achieved. The constructive proof under an additional assumption on the weights γ_i is given in [33]. This is achieved by the *WTP* algorithm which can be applied not only to multivariate integration but also to multivariate approximation for weighted tensor product spaces.

Motivated by strong tractability results for multivariate integration and approximation in the worst case setting, we try to establish similar results for the quasilinear contraction. We now briefly outline the main results of this paper. Strong tractability of any problem can hold only for certain (generally compact) subspaces F_d of the space B_d of functions of d variables. We show that strong tractability of the conditional expectation problem in the space F_d implies strong tractability of the quasilinear contraction problem as along as the successive approximation V_i and the solution V^* belong to the same space F_d. As already mentioned the conditional expectation problem corresponds to weighted multivariate integration. To check strong tractability of the conditional expectation problem we consider two cases depending on Markov transition densities $p_k(t|s)$.

The first simplified case is when these densities do not depend on the second argument

s. Then it is relatively easy to show sufficient conditions on strong tractability of the conditional expectation problem. If F_d is a Hilbert space with reproducing kernel K_d then strong tractability holds for any densities p_k if, in particular, the values of $K_d(x, x)$ are uniformly bounded in x and d. Furthermore, the strong exponent is at most 2. The latter assumption is satisfied for weighted Sobolev spaces.

The second case is for general Markov transition densities which may depend on the second argument. In this case, we show that strong tractability of approximation in F_d implies strong tractability of the conditional expectation problem and also strong tractability of the quasilinear problem as long as the approximation V_i and the solution V^* belong to the same space F_d. We check that these assumptions hold in a weighted Sobolev space.

Section 2 introduces the class of quasilinear fixed points and shows how this class nests a number of important applications in the economics and engineering literatures. Section 3 introduces a class of deterministic algorithms for computing the quasilinear fixed points. Section 4 provides sufficient conditions for strong tractability of the quasilinear contraction problem for constant Markov transition densities whereas Section 5 does it for general Markov transition densities.

2 Quasilinear Contraction Mappings

In this section we define a quasilinear contraction mapping as a nonlinear operator of a finite number of linear conditional expectations. We use the adjective "quasilinear" since the nonlinear operator defining the contraction mapping satisfies a restricted form of linearity given in Definition 1 below.

Definition 1: We say a function $f : \mathbb{R}^m \to \mathbb{R}$ is *quasilinear* if:

1. f is continuous and nondecreasing in each argument,
2. For all $\beta \in R$ we have $f(x + \beta\vec{e}) = f(x) + \beta$ where $\vec{e} = [1, 1, \ldots, 1] \in \mathbb{R}^m$. $\square$

We present several examples of quasilinear functions f.

1. *Linear functions.* Let f be given by

$$f(x) = \sum_{k=1}^{m} c_k x_k, \tag{2}$$

where $c_k \geq 0$, and $\sum_{k=1}^{m} c_k = 1$.

2. *Max function.* Let f be given by

$$f(x) \;=\; \max_{k=1,2,\ldots,m} x_k. \tag{3}$$

3. *Smoothed Max function.* Let f_σ be given by

$$f_\sigma(x) \;=\; \sigma \ln \left(\sum_{k=1}^{m} \exp\left(x_k/\sigma\right) \right) \tag{4}$$

for a nonnegative σ. Observe that

$$0 \;\leq\; f_\sigma(x) - \max_{k=1,2,\ldots,m} x_k \;\leq \sigma \ln m, \tag{5}$$

and therefore $\lim_{\sigma \to 0} f_\sigma(x) = \max_{1 \leq k \leq m} x_k$.

4. *"Social Surplus functions".* Let f_σ be given by

$$f_\sigma(x) = \int_{\mathbb{R}^m} \max_{k=1,2,\ldots,m} \left[x_k + \sigma\xi_k\right] q(\xi_1, \ldots, \xi_m) d\xi_1 \cdots d\xi_m, \tag{6}$$

where $q(\xi_1, \ldots, \xi_m)$ is a probability density over $\mathbb{R}^m$ which has finite absolute first moments. This class of functions was introduced in [14]. The adjective "Social Surplus" reflects the interpretation of $f_\sigma(x)$ as the expected utility of a population of agents indexed by $\xi = (\xi_1, \ldots, \xi_m)$ who face m possible choices, where the utility of choice k is $x_k + \sigma\xi_k$. The smoothed max function (4) is a special case of (6) when q is the product of m appropriately standardized univariate Type III extreme value distributions, i.e.

$$q(\xi_1, \ldots, \xi_m) = \prod_{i=1}^{m} \exp\{-(\xi_i + \gamma)\} \exp\left\{-\exp\{-(\xi_i + \gamma)\}\right\}$$

where $\gamma \simeq .577216\ldots$ is Euler's constant.

As noted in section 1, B_d denotes the Banach space of continuous functions V on a compact set $S_d \subset \mathbb{R}^d$ with nonempty interior equipped with the sup-norm, $\|V\| = \sup_{s \in S_d} |V(s)|$. Let

$$u_k : S_d \to \mathbb{R} \qquad \text{for} \qquad k = 1, 2, \ldots, m \tag{7}$$

be elements of B_d, and let

$$p_k : S_d \times S_d \to \mathbb{R}_+ \qquad \text{for} \qquad k = 1, 2, \ldots, m \tag{8}$$

be Markov transition densities which are weakly continuous in their second argument. That is, $p_k(\cdot|s)$ is a probability density for each $s \in S_d$ and has the property that for each $V \in B_d$ we have $E_k V \in B_d$, where the conditional expectation operator E_k is given by

$$E_k V(s) = \int_{S_d} V(t) p_k(t|s)\, dt. \tag{9}$$

Definition 2: We say a mapping $\Gamma : B_d \to B_d$ is *quasilinear* if Γ is given by:

$$\Gamma(V)(s) = f(u_1(s) + \beta E_1 V(s), \ldots, u_m(s) + \beta E_m V(s)), \tag{10}$$

where $f : R^m \to R$ is a quasilinear function, and the constant (discount factor) β belongs to $(0, 1)$.

We now present several examples of quasilinear mappings.

1. *Contractive Fredholm Integral Equations*

   When $m = 1$ and $f(x) = x$ we have

   $$\Gamma(V)(s) \,=\, u(s) \,+\, \beta\, EV(s), \qquad \forall\, s \in S_d. \tag{11}$$

   Thus Γ is a linear operator in this case. It is easy to see that the fixed point problem $V = u + \beta EV$ is equivalent to solving a Fredholm equation of the second kind with kernel $k(t, s) = \beta p(t|s)$. Equations of this form arise in rational expectations theories of asset pricing. See, for example, [13], or [28]. A lot is known about the complexity of computing approximate solutions to Fredholm integral equations with general kernels, see [34] and [10] for surveys of deterministic and stochastic complexity bounds for this problem. In the case of general kernels, the results in [34] show that the problem is intractable in the the worst case using deterministic algorithms. By exploiting the special structure of the kernel for the class of contractive Fredholm integral equations combined with additional special structure on the functions u and p (to be defined shortly) we will be able to show that the fixed point problem, and the associated Fredholm problem, is strongly tractable.

2. *Bellman operators.*

   We take $f(x) = \max_{1 \le k \le m} x_k$ and obtain

   $$\Gamma(V)(s) = \max_{k=1,2,\ldots,m} \left[u_k(s) + \beta E_k V(s)\right].$$

   The associated fixed point equation, $V = \Gamma(V)$ is known as *Bellman's equation,* the fundamental equation underlying infinite horizon Markovian decision problems (see, e.g. [4] and [7]). Bellman operators pose difficulties for our analysis since our results depend on the ability to exploit smoothness properties of the function $\Gamma(V)$. However even if the functions u_k and p_k are very smooth with respect to s, the function $\Gamma(V)$ will generally only be a Lipschitz continuous function of s due to the presence of "kinks" induced by the max operator.

3. *Smoothed Bellman Operators.*

   We take $f_\sigma(x) = \sigma \ln(\sum_{k=1}^m \exp(x_k/\sigma))$ and obtain

   $$\Gamma_\sigma(V)(s) = \sigma \ln \left[\sum_{k=1}^m \exp\left[\frac{1}{\sigma}[u_k(s) + \beta E_k V(s)]\right]\right]. \tag{12}$$

   The function $\Gamma(V)$ is as smooth as u_k and p_k. From (5) we conclude

   $$0 \le \Gamma_\sigma(V)(s) - \Gamma(V)(s) \le \sigma \ln m, \qquad \forall\, s \in S_d,$$

   and therefore the Bellman operator is a uniform limit of smoothed Bellman operators: $\Gamma = \lim_{\sigma \to 0} \Gamma_\sigma$. This justifies the name of Γ_σ as the smoothed Bellman operator. Fixed points problems with smoothed Bellman operators arise in econometric applications, see e.g., [23].

4. *Smoothed Bellman Operators via Social Surplus Functions.*

   A wider class of smoothed Bellman operators can be defined for the class of social surplus functions f_σ given in equation (6).

   $$\Gamma_\sigma(V)(s) = \int_{\mathbb{R}^m} \max_{k=1,\ldots,m} [u_k(s) + \beta E_k V(s) + \sigma \xi_k] q(\xi_1, \ldots, \xi_m) d\xi_1 \cdots d\xi_m. \tag{13}$$

   It is obvious from (13) that $\Gamma = \lim_{\sigma \to 0} \Gamma_\sigma$. While our analysis of the complexity of the quasilinear fixed point problem explicitly considers the numerical integration problem underlying the evaluation of the conditional expectation operators, $\{E_k\}$, it abstracts from the integration problem defining the quasilinear function in (13). Thus, we will assume that the function f in equation (10) can be evaluated exactly, such as in the case (12) (abstracting from potential errors in approximating $\exp\{x\}$ and $\log\{x\}$ which we presume are of second order relative to errors in multivariate integration). Otherwise the analysis becomes even more complicated since we need to control approximation error in the quasilinear function f in addition to the approximation error in $\{E_k\}$.

Theorem 1 *Let Γ be a quasilinear contraction mapping given in Definition 2. Then Γ is a contraction mapping.*

Proof: . We have:

$$\Gamma(V)(s) = \Gamma(W + V - W)(s) \leq \Gamma(W + \|V - W\|\vec{e})(s) = \Gamma(W)(s) + \beta\|V - W\|,$$

using the properties 1 and 2 in Definition 1 above. Repeating the same argument, but interchanging V and W we get

$$|\Gamma(V)(s) - \Gamma(W)(s)| \leq \beta\|V - W\|, \qquad \forall\, s \in S_d,$$

i.e., Γ is a contraction mapping. □

While Definition 2 encompasses a fairly broad class of problems of interest (including dynamic programming, asset pricing and econometric problems), there are problems of interest that do not fall within this framework. An example is provided by problems encountered in recursive utility theory. This theory (see [8] for an introduction), provides recursive representations for preferences that may be non time-separable or may not satisfy the expected utility hypothesis. Similar objective functions emerge from the engineering literature on robust or risk-sensitive control theory. In a recent paper, Anderson, Hansen and Sargent [1] showed that the following functional equation could be derived from either of these frameworks:

$$V(s) = u(s) + \frac{1}{\sigma}\log\left[E\exp\{\sigma V\}(s)\right]. \tag{14}$$

This functional equation does not fall within Definition 2 since the operator is defined in terms of conditional expectations of nonlinear functions of V instead of V directly. Despite this nonlinearity, it is easy to show that the operator $\Gamma_\sigma(W)$ defined by

$$\Gamma_\sigma(W)(s) = \frac{1}{\sigma}\log\left[E\exp\{\sigma V\}(s)\right], \tag{15}$$

is also a contraction mapping. We believe the results in this paper can be extended to these more general types of quasilinear contraction mappings, but to simplify our arguments and avoid technical complications we will focus on the restricted class of quasilinear contractions given in Definition 1.

Quasilinear mappings simplify if all Markov transition densities are the same, $p_k \equiv p$. Then the second property of the quasilinear function f yields

$$\Gamma(V)(s) \;=\; \pi(s) \;+\; \beta \int_{S_d} V(t)p(t|s)\,dt, \quad \text{with } \pi(s) \;=\; f\left(u_1(s), \ldots, u_m(s)\right). \tag{16}$$

If we further assume that the Markov transition density does not depend on s, $p(t|s) = p(t)$, then

$$\Gamma(V)(s) \;=\; \pi(s) \;+\; \beta \int_{S_d} V(t)p(t)\,dt. \tag{17}$$

In the later case, the fixed point $V = \Gamma(V)$ differs from π only by a constant, i.e.

$$V(s) = \pi(s) + \frac{\beta}{1-\beta} \int_{S_d} \pi(t)p(t)dt. \tag{18}$$

Hence, computation of V reduces to the computation of a single multivariate integral in this case.

The *quasilinear contraction problem* is defined as the problem of computing an approximation of $V^*(s)$ to within $\varepsilon \in (0,1)$. Here s is a given point from S_d and V^* is the unique solution to the contraction fixed point problem

$$V = \Gamma(V), \tag{19}$$

and Γ is a quasilinear contraction mapping satisfying Definition 2. The existence and uniqueness of V^* follows from the Banach fixed theorem.

More precisely, we want to compute an ε-approximation $\hat{V}(s)$ such that

$$|V^*(s) - \hat{V}(s)| \;\le\; \varepsilon \, \max\left(\|V^*\|, |||V^*||| \right). \tag{20}$$

Here $|||\cdot|||$ is a norm which maybe be different than the usual sup norm $\|\cdot\|$ on the space B_d. As we shall see later, the choice of the norm $|||\cdot|||$ is very important and our results on the strong tractability of the quasilinear contraction problem do depend on this norm. The only restriction on the norm $|||\cdot|||$ is that $|||V^*|||$ is well defined. Note that if we choose $|||\cdot|||$ such that $|||V^*|||$ is much larger than $\|V^*\|$ then the problem of computing an ε-approximation is easier.

We comment on our assumption that we want to approximate the contraction fixed point only at a one point. Obviously, if we want to compute such approximations at a number of points $s_1, s_2, \ldots, s_k$, it is enough to apply our results k times. As long as k is relatively small it can be done without an essential increase of the total cost. If, however, k is large this approach is questionable and in many cases it will be better to approximate the contraction fixed point V^* as a function and then use such an approximation at all points s_i. The latter problem is not considered in this paper.

3 Algorithms

In this section we consider various algorithms for solving the quasilinear contraction problem. We begin by assuming for a moment that we can exactly evaluate $\Gamma(W)(s)$ for a given function W. Then we can solve (19) by the simple iteration

$$V_i(s) \,=\, \Gamma(V_{i-1})(s), \qquad i = 1, 2\ldots, \tag{21}$$

where V_0 is the initial approximation of the solution V. For simplicity we take $V_0 = 0$. Clearly,

$$\|V_i - V^*\| \,\le\, \beta^i \, \|V^*\|, \qquad \forall\, i.$$

Let

$$\varepsilon_1 \,=\, \varepsilon \, \frac{\max\left(\|V^*\|, |||V^*||| \right)}{\|V^*\|}. \tag{22}$$

We compute an ε-approximation $A(s) = V_n(s)$ if $\beta^n \le \varepsilon_1$ which holds for

$$n \,=\, \left\lceil \frac{\ln\, 1/\varepsilon_1}{\ln\, 1/\beta} \right\rceil. \tag{23}$$

Observe that the formula for the number of steps n is formally *not* constructive since it depends on the norms of the unknown solution V^*. However, we have $\varepsilon_1 \ge \varepsilon$ and we may bound n by replacing ε_1 by ε, $n \le \lceil \ln(1/\varepsilon)/\ln(1/\beta) \rceil$. The latter bound is constructive.

If β is not too close to 1, the number n of steps is quite reasonable. In this paper we assume that this is indeed the case. A number of estimates presented in this paper have unspecified factors which depend on β. These factors are of order 1 if β is not too close to 1. The case of β close to 1 is also of interest although it is not studied in this paper. For β close to 1, the iteration (21) as well as all its modifications studied in this paper can be significantly improved for moderate values of d and ε^{-1}, and the number of steps can be proportional to $\ln 1/(1-\beta)$ as shown in [12, 27].

Thus, as long as $\Gamma(V_n)(s)$ can be computed exactly, the quasilinear contraction problem can be solved quite efficiently. However, the assumption on the exact computation of $\Gamma(V_n)(s)$ is not realistic. Indeed, the computation of $\Gamma(V_n)(s)$ requires in particular the computation of the d-dimensional integrals with the weights $p_k(\cdot|s)$. This can be done, in general, only approximately.

Assume then that instead of (21) we can compute the perturbed sequence

$$V_i(s) \;=\; \hat{\Gamma}(V_{i-1}) \;=\; \Gamma(V_{i-1})(s) + \delta_{i-1}(s), \qquad \text{with } |\delta_{i-1}(s)| \le \delta_{i-1}\|V_{i-1}\|, \tag{24}$$

for some nonnegative δ_{i-1}. We will see later that δ_{i-1} corresponds to the quadrature error and can be made sufficiently small by taking sufficiently many sample points in the quadrature formula.

It is natural to ask how small δ_{i-1} should be to preserve the global convergence property of the sequence (21). In what follows, we assume that

$$\beta + \delta_{i-1} \;\le\; \overline{\beta} < 1, \qquad \forall\, i = 1, 2, \ldots . \tag{25}$$

We have

$$\begin{aligned} \|V_i - V^*\| \;&\le\; \beta\,\|V_{i-1} - V^*\| + \delta_{i-1}\|V_{i-1}\| \;\le\; (\beta + \delta_{i-1})\|V_{i-1} - V^*\| + \delta_{i-1}\|V^*\| \\ &\le\; \overline{\beta}\,\|V_{i-1} - V^*\| + \delta_{i-1}\|V^*\|. \end{aligned}$$

This yields

$$\|V_i - V^*\| \;\le\; \overline{\beta}^{\,i}\,\|V^*\| + \left(\sum_{j=0}^{i-1} \overline{\beta}^{\,i-1-j}\delta_j\right)\|V^*\|.$$

If we set

$$\overline{\beta}^{\,n} \;\le\; \varepsilon_1/2 \qquad \text{which holds for} \qquad n \;=\; \left\lceil \frac{\ln 2/\varepsilon_1}{\ln 1/\overline{\beta}} \right\rceil,$$

and

$$\overline{\beta}^{\,n-1-i}\delta_i \;\le\; \frac{\varepsilon_1}{2n}, \qquad i = 0, 1, \ldots, n-1,$$

then $A(s) = V_n(s)$ is an ε-approximation since

$$\|V_n - V^*\| \le \varepsilon \max(\|V^*\|, |||V^*|||).$$

Observe that the number n of steps is still reasonable if we choose $\overline{\beta}$ close to β, and β is not too close to 1. The perturbation parameters δ_i can be defined as

$$\delta_i \le \frac{\overline{\beta}^{\,i} \ln 1/\overline{\beta}}{\ln 2/\varepsilon_1 + \ln 1/\overline{\beta}} \le \frac{\varepsilon_1}{2n\overline{\beta}^{\,n}}\,\overline{\beta}^{\,i+1}. \tag{26}$$

Hence, δ_i may mildly depend on ε_1, and should decrease geometrically with i. This means that we need more accuracy as we go along, and this is quite natural.

Knowing how much we can perturb the original simple iteration (21) we are ready to replace multivariate integrals in $\Gamma(V)$ by quadrature formulas. We approximate the conditional expectation E_k given by (9) by quadrature formulas

$$\hat{E}_{k,j}(V)(s) = \sum_{i=1}^{j} a_{i,j,k}(s) V(t_{i,j,k}) \tag{27}$$

which use the j function values V. Here, $a_{i,j,k}(s)$ are real numbers and $t_{i,j,k}$ are sample points from S_d. For instance, we can take $a_{i,j,k}(s) = p_k(s_i|s)/j$ and $t_{i,j,k} = s_i$ for some sample points s_i. For $S_d = [0,1]^d$ we can take s_i as independent random points which are uniformly distributed over $[0,1]^d$. In this case, $\hat{E}_{k,j}$ corresponds to the classical Monte Carlo algorithm. We may also take s_i as low discrepancy deterministic sample points. In this case, $\hat{E}_{k,j}$ corresponds to a quasi-Monte Carlo algorithm.

The quality of the quadrature formula $\hat{E}_{k,j}$ will be measured by its error. We assume that $\varepsilon_j(V)$ is an upper bound on the quadrature error,

$$\left| E_k(V)(s) - \hat{E}_{k,j}(V)(s) \right| \le \varepsilon_j(V), \qquad \forall\, s \in S_d,\ k = 1, 2, \ldots, m. \tag{28}$$

We are ready to modify the algorithm (21) by replacing the weighted integrals $E_k(V_i)(s)$ by the quadratures $\hat{E}_{k,j_i}(V_i)(s)$. Here, the not yet specified sequence $\{j_i\}$ tells us how many sample points are used in the quadrature formulas. The choice of $\{j_i\}$ will depend on the errors $\varepsilon_j(V)$.

The modified sequence (21) is now formally given by

$$V_i(s) \;=\; f\left(u_1(s) + \beta \hat{E}_{1,j_{i-1}}(V_{i-1})(s), \dots, u_m(s) + \beta \hat{E}_{m,j_{i-1}}(V_{i-1})(s)\right), \quad \forall\, i = 1, 2, \dots\,. \tag{29}$$

We are ready to prove.

Theorem 2 *Assume that the quadrature errors satisfy*

$$\varepsilon_{j_i}(V_i) \;\le\; \min\left(\frac{1-\beta}{2}, \frac{\left(\frac{1+\beta}{2}\right)^i \ln \frac{2}{1+\beta}}{\beta\left(\ln \frac{2}{\varepsilon_1} + \ln \frac{2}{1+\beta}\right)}\right) \|V_i\| \tag{30}$$

for $i = 1, 2, \dots$ *and* $n = \lceil (\ln 2/\varepsilon_1)/(\ln 2/(1+\beta)) \rceil$. *Then* $A(s) = V_n(s)$ *given by (29) is an* ε*-approximation to* $V^*(s)$.

Proof: It is easy to check that (29) is of the form (24) with

$$\delta_{i-1} \;=\; \beta\, \frac{\varepsilon_{j_{i-1}}(V_{i-1})}{\|V_{i-1}\|}.$$

Let $\overline{\beta} = (1+\beta)/2$. From (30) we conclude that (25) as well as (26) hold and therefore $A(s)$ is an ε-approximation. □

We now discuss the cost of the algorithm $A(s) = V_n(s)$ given by (29) and Theorem 2 with

$$n \;=\; \left\lceil \frac{\ln 2/\varepsilon_1}{\ln 2/(1+\beta)} \right\rceil. \tag{31}$$

We assume that we can compute the values of the functions u_k and f at any point as well as we can perform arithmetic operations. Let the cost of one evaluation of the functions u_k be $c(u)$ and let the cost of one evaluation of the function f be $c(f)$. Observe that u_k is a function of d variables and therefore the cost $c(u)$ may depend on d. Similarly, f is a function of m variables and therefore the cost $c(f)$ may depend on m. We assume that the cost of one arithmetic operation is taken as unity. We also assume that the sample points $t_{i,j_i,k}$ as well $\beta\, a_{i,j_i,k}(s)$ are "precomputed". Usually, these precomputed numbers depend on the Markov transition densities $p_k(\cdot|s)$ and may require a number of evaluations of $p_k(\cdot|s)$. Since this should be done once for a given ε we do not include the cost of generating of these precomputed numbers as is typically assumed in the complexity analysis, see [30] and [17] where this point is fully discussed.

We now explain in detail how $A(s)$ can be computed. For $i = 1, 2, \ldots, n$ we use

$$V_i(s) = f\Big(u_1(s) + \sum_{p=1}^{j_{i-1}} \beta a_{p,j_{i-1},1}(s) V_{i-1}(t_{p,j_{i-1},1}), \ldots, u_m(s) + \sum_{p=1}^{j_{i-1}} \beta a_{p,j_{i-1},m}(s) V_{i-1}(t_{p,j_{i-1},m})\Big). \tag{32}$$

Observe that for $i = 1$ we have $V_0 = 0$ and the sample points $t_{p,j_0,k}$ are not needed. This corresponds formally to $j_0 = 0$. Let

$$T_{\varepsilon,d} \;=\; \Big\{\, t_{p,j_{i-1},k} \;:\; i \in [2,n],\, p \in [1, j_{i-1}],\, k \in [1,m] \,\Big\} \tag{33}$$

denote all integration sample points used in (32), and let $|T_{\varepsilon,d}|$ denote the cardinality of the set $T_{\varepsilon,d}$. Clearly,

$$|T_{\varepsilon,d}| \;\le\; m\ (j_1 + j_2 + \cdots + j_{n-1})\,. \tag{34}$$

Furthermore, if we use *nested* integration sample points, $\{t_{p,j_{i-1},k}\} \subset \{t_{p,j_i,k}\}$ then

$$|T_{\varepsilon,d}| \;\le\; m\ j_{n-1}.$$

If the points $\{t_{p,j,k}\}$ are the same for all $k = 1, \ldots, m$, then

$$|T_{\varepsilon,d}| \;\le\; j_{n-1}.$$

In order to compute $A(s) = V_n(s)$ we need to compute V_{n-1} at the sample points $t_{p,j_{n-1},k}$. This can be achieved if we know V_{n-2} at the sample points $t_{p,j_{n-2},k}$ and so on. Therefore we compute successively $V_1, V_2, \ldots, V_{n-1}$ at *all* the sample points of the set $T_{\varepsilon,d}$, and then we can compute the value $V_n(s)$.

More specifically, we first $u_k(x)$ for all $x \in T_{\varepsilon,d} \cup \{s\}$ and $k \in [1,m]$ at cost $m(1 + |T_{\varepsilon,d}|)\, c(u)$. To compute $V_i(x)$ for all $x \in T_{\varepsilon,d}$, we perform $2 j_{i-1} m$ arithmetic operations and one evaluation of the function f at cost $(2 j_{i-1} m + c(f))|T_{\varepsilon,d}|$. Finally we compute $V_n(s)$ at cost $2 j_{n-1} m + c(f)$. Then cost(A) of computing $A(s)$ is

$$\mathrm{cost}(A) \;=\; m(1 + |T_{\varepsilon,d}|)\, c(u) \;+\; (1 + n|T_{\varepsilon,d}|) c(f) \;+\; 2m|T_{\varepsilon,d}| \sum_{i=1}^{n} j_{i-1} \;+\; 2 j_{n-1} m$$

This means that the cost of the algorithm A crucially depends on the cardinality of $T_{\varepsilon,d}$ and on the values of j_i which, in turn, depend on the efficiency of quadrature formulas (27). More precisely, the cost of A depends for which indices j we can guarantee that the integration error $\varepsilon_j(V)$ satisfies (30). Note that (30) holds if we set

$$\varepsilon_{j_i}(V_i) \;=\; C_\beta \left(\frac{1+\beta}{2}\right)^i \frac{\|V_i\|}{\ln \varepsilon_1^{-1}} \tag{35}$$

for some C_β depending only on β. Let

$$N(\varepsilon, d) \;=\; j_1 + j_2 + \cdots + j_{n-1} \tag{36}$$

denote the total number of integration steps needed to computed $A(s)$. Then the cost of the algorithm A can be rewritten as

$$\mathrm{cost}(A) \;=\; O\left(N(\varepsilon, d)\, \left(m\, c(u) \;+\; (\ln \varepsilon^{-1})\, c(f)\right) \;+\; N^2(\varepsilon, d)\, m \right) \tag{37}$$

with the big O-factor depending only on β.

The essence of (37) is that the cost of the algorithm A depends polynomially on $N(\varepsilon, d)$. In fact, the dependence is roughly linear in $N(\varepsilon, d)$ in terms of the cost of $c(u)$ and $c(f)$ and quadratic in $N(\varepsilon, d)$ in terms of the cost of arithmetic operations. In many cases, the cost $c(u)$ or $c(f)$ is much larger than unity and therefore the total cost can be proportional to $N(\varepsilon, d)$. Hence, as long as $N(\varepsilon, d)$ is not too large the cost of the algorithm A is reasonable.

The quantity $N(\varepsilon, d)$ measures the difficulty of approximating the conditional expectations E_k . From the perspective of computing high-dimensional fixed points where d is large, the best possible case is where $N(\varepsilon, d)$ can be bounded by a polynomial in ε^{-1} independently[1] of d.

At this point we need to recall the concept of strong tractability of computational problems, see [35]. Suppose we want to compute an ε-approximation of $E_k(V)(s)$ for all $s \in S_d$ and for all functions V from a normed space F_d which is a subset of B_d. Let $|||V||| = \|V\|_{F_d}$ be the norm of the space F_d. Let $n = n(\varepsilon, d)$ be the smallest integer for which there exists $\hat{E}_{k,n}$ of the form (27) such that

$$|E_k(V)(s) - \hat{E}_{k,n}(V)(s)| \;\le\; \varepsilon\, \|V\|_{F_d}, \qquad \forall\, s \in S_d,\, V \in F_d,\, k \in [1, m].$$

---

[1] It is also reasonable to study the case of *tractability* in which we also permit a polynomial dependence on d. For simplicity the focus of our attention in this paper is on strong tractability.

We say that the conditional expectation problem is *strongly tractable* in F_d if there exist nonnegative C and p such that

$$n(\varepsilon,d) \,\le\, C\,\varepsilon^{-p}, \qquad \forall\,\varepsilon\in(0,1),\; d=1,2,\dots\,. \tag{38}$$

The smallest (or infimum of) exponent p in the latter bound is called the the *strong exponent* of the conditional expectation problem.

It seems natural to extend this definition to the quasilinear contraction problem. We say that the quasilinear contraction problem is *strongly tractable* iff there exist nonnegative numbers C, p and p_1 such that the cost of computing an ε-approximation can be bounded by

$$C\,\left([c(u)+c(f)]\,\varepsilon^{-p}\,+\,\varepsilon^{-p_1}\right).$$

Hence, p is the exponent of ε^{-1} which tells us how many evaluations of u and f are needed, and p_1 is the exponent of ε^{-1} which tells how many arithmetic operations are needed to solve the quasilinear contraction problem. The smallest (or infima of) exponents p and p_1 are called the *strong exponent of information* and the *strong exponent of arithmetic operations* of the quasilinear contraction problem.

It is easy to check that strong tractability of the conditional expectation problem in F_d implies strong tractability of the quasilinear contraction problem as long as V_i and the solution V^* of the quasilinear contraction problem belong to the space F_d. Indeed, define

$$j_i \,=\, C\,\left(C_\beta\,\frac{\|V_i\|}{\|V_i\|_{F_d}}\,\left(\frac{1+\beta}{2}\right)^i\,\frac{1}{\ln\,\varepsilon_1^{-1}}\right)^{-p} \tag{39}$$

with C and p given by (38), and C_β by (35). Then there exists $\hat{E}_{k,j_i}$ of the form (27) such that

$$|E_k(V_i)(s)-\hat{E}_{k,j_i}(V_i)(s)| \,\le\, C_\beta\,\left(\frac{1+\beta}{2}\right)^i\,\frac{\|V_i\|}{\ln\,\varepsilon_1^{-1}}.$$

Hence (35) is satisfied. We now estimate $N(\varepsilon,d)$ for j_i given by (39). Then

$$N(\varepsilon,d) \,=\, O\left(C\,\left(\max_{i\in[1,n]}\frac{\|V_i\|_{F_d}}{\|V_i\|}\right)^p\,\left(\varepsilon_1^{-1}\,\ln\,\varepsilon_1^{-1}\right)^p\right).$$

Using the definition of ε_1, see (22), we have

$$N(\varepsilon,d) \,=\, O\left(C\,\left(\max_{i\in[1,n]}\frac{\|V_i\|_{F_d}}{\|V_i\|}\,\frac{\|V^*\|}{\|V^*\|_{F_d}}\right)^p\,\left(\varepsilon^{-1}\,\ln\,\varepsilon^{-1}\right)^p\right)$$

with the big O-factor depending only on β. Note that $\|V_i - V^*\| \le \beta^i \|V^*\|$ implies that $\|V_i\| \ge (1-\beta^i)\|V^*\| \ge (1-\beta)\|V^*\|$. Hence $\|V^*\|/\|V_i\| \le 1/(1-\beta)$ and we can drop this ratio from the last maximum at the expense of enlarging the factor in the big O notation.

We summarize this as well as the previous analysis in the following theorem.

Theorem 3 *The algorithm A given by (29) and (31) computes an ε-approximation to the solution of the quasilinear contraction problem at cost*

$$\mathrm{cost}(A) \,=\, O\left(c(u)\, m\, N(\varepsilon,d) \,+\, c(f)\, N(\varepsilon,d)\, \ln \varepsilon^{-1} \,+\, m\, N^2(\varepsilon,d)\right)$$

where $N(\varepsilon,d)$ is given by (36). Suppose that

$$M \,=\, \sup_{d,i=1,2,\dots} \frac{\|V_i\|_{F_d}}{\|V^*\|_{F_d}} \tag{40}$$

is finite. Then strong tractability of the conditional expectation problem in F_d with (38) implies strong tractability of the quasilinear contraction problem and the cost of the algorithm A satisfies

$$\mathrm{cost}(A) = O\left(c(u)\, m\, C\, M^p\, \varepsilon^{-p}\, \ln^p \varepsilon^{-1} + c(f)\, C\, M^p\, \varepsilon^{-p}\, \ln^{p+1} \varepsilon^{-1} + m\, C^2\, M^{2p}\, \varepsilon^{-2p}\, \ln^{2p} \varepsilon^{-1}\right)$$

with the big O-factor depending only on β.

Hence, the strong exponent of the quasilinear contraction problem is at most equal to the strong exponent of the conditional expectation problem whereas the strong exponent of arithmetic operations of the quasilinear contraction problem is at most twice as the strong exponent of the conditional expectation problem.

Theorem 3 relates strong tractability of the conditional expectation problem and the quasilinear contraction problem as long as we know that V_i and V are in the space F_d and the bound M of (40) is finite. So far we only know that V_i and V^* belong to the Banach space B_d of continuous functions. Let us then take $F_d = B_d$. Mere continuity of functions in B_d is not only enough to guarantee tractability of the conditional expectation problem but is not even enough to guarantee solvability of the conditional expectation problem. The reason is that integration errors do not go to zero for the class of continuous functions, see e.g. [30]. If we want to have the space B_d, one possible solution is to switch to the randomized setting in which the bound (28) is understood as the expected error with respect to randomized sample points. This approach is used in [24]. In this paper we stay with the worst case setting and deterministic algorithms. Therefore to obtain strong tractability we must explore additional properties of the quasilinear contraction problem so that the approximations V_i and the solution V^* belong to spaces F_d for which the conditional expectation problem is strongly tractable.

4 The *IID* Case

It is easiest to explain our results by beginning with the special case of Markov transition densities $p_k(t|s) = p_k(t)$ which are independent of the second argument s. This implies that the realizations from these densities, $\{s_{i,k}\}$, $i = 1, 2, \dots$ are independent and identically distributed (*IID*) sequences. The general Markov case, where transition densities are allowed to depend on s, is considered in the next section. For the *IID* case it is easier to specify the spaces F_d for which the conditional expectation problem as well as the quasilinear contraction problem are strongly tractable.

For the densities p_k independent of s, the assumption on the quality of the quadrature rules (28) simplifies. We may, of course, assume now that $a_{i,j,k}(s) = a_{i,j,k}$ is also independent of s, and we have

$$\left| \int_{S_d} V(t) p_k(t)\, dt \;-\; \sum_{i=1}^{n} a_{i,j,k} V(t_{i,j,k}) \right| \;\le\; \varepsilon_j(V) \qquad k = 1, 2, \dots, m. \tag{41}$$

As in the previous section we take $|||\cdot||| = \|\cdot\|_{F_d}$ for some normed space F_d. We assume that F_d is a Hilbert space with reproducing kernel $K_d : S_d \times S_d \to R$. For the basic theory of such spaces the reader is referred to [2, 32]. The inner product and norm in F_d are denoted by $\langle \cdot, \cdot \rangle$ and $|||\cdot||| = \|\cdot\|_{F_d} = \langle \cdot, \cdot \rangle^{1/2}$. For $V \in F_d$ we have $V(t) = \langle V, K_d(\cdot, t) \rangle$. For $k = 1, 2, \dots, m$, define

$$h_k(x) \;=\; \int_{S_d} K_d(x, t) p_k(t)\, dt, \qquad x \in S_d. \tag{42}$$

Assuming that $h_k \in F_d$ we have from (9) and (27)

$$E_k(V) \;=\; \langle V, h_k \rangle\,, \qquad \hat{E}_{k,j}(V) \;=\; \left\langle V, \sum_{i=1}^{j} a_{i,j,k} K_d(\cdot, t_{i,j,k}) \right\rangle.$$

It is easy to check that

$$\|E_k\|^2 \;=\; \|h_k\|_{F_d}^2 \;=\; \int_{S_d} h_k(x)p_k(x)\,dx \;=\; \int_{S_d\times S_d} K_d(x,t)p_k(x)p_k(t)\,dt\,dx.$$

¿From this we see that (41) holds with

$$\varepsilon_j(V) \;=\; \|V\|_{F_d} \max_{k\in[1,m]} \left\| h_k - \sum_{i=1}^{j} a_{i,j,k} K_d(\cdot, t_{i,j,k}) \right\|_{F_d}.$$

We now take $a_{i,j,k}(s) = 1/j$. We have

$$\left\| h_k - \frac{1}{j}\sum_{i=1}^{j} K_d(\cdot, t_{i,j,k}) \right\|_{F_d}^2 \;=\; \|h_k\|^2 - \frac{2}{j}\sum_{i=1}^{j} h_k(t_{i,j,k}) + \frac{1}{j^2}\sum_{i,l=1}^{j} K_d(t_{i,j,k}, t_{l,j,k}).$$

Consider the sample points $t_{i,j,k}$ which are *iid* draws from the measure with the density p_k. Integrating over such sample points we get

$$E\left(\left\| h_k - \frac{1}{j}\sum_{i=1}^{j} K_d(\cdot, t_{i,j,k}) \right\|_{F_d}^2 \right) \;=\; \frac{\rho_{d,k}}{j},$$

where

$$\rho_{d,k} \;=\; \int_{S_d} K_d(x,x)p_k(x)\,dx - \int_{S_d^2} K_d(x,t)p_k(x)p_k(t)\,dx\,dt. \tag{43}$$

¿From the mean value theorem we conclude that there exist sample points $t_{i,j,k}$ such that the quadrature formula $\hat{E}_{k,j}$ of the form (27) satisfies

$$\left| E_k(V) - \hat{E}_{k,j}(V) \right| \;\le\; \|V\|_{F_d} \frac{\sqrt{\rho_{d,k}}}{\sqrt{j}}. \tag{44}$$

We stress that the proof of (44) is *non-constructive* since we use the mean value theorem. This implies that the algorithm A with $\hat{E}_{k,j}$ is also non-constructive.

The estimate (44) proves that (38) holds with $p = 2$ and

$$C \;=\; \sup_{k\in[1,m],\, d=1,2,\dots} (1 + \rho_{d,k}) \tag{45}$$

as long as C is finite. We thus have the following theorem.

Theorem 4 *Suppose that C given by (45) is finite. Then the conditional expectation problem is strongly tractable with strong exponent at most 2 Suppose additionally that M given by (40) is finite. Then the quasilinear contraction problem is strongly tractable with strong exponent of information at most 2 and strong exponent of arithmetic operations at most 4. Furthermore, the algorithm A defined by (29) with (non-constructive) $\hat{E}_{k,j}$ computes an ε-approximation at cost*

$$\mathrm{cost}(A) \,=\, O\left(c(u)\, m\, C\, M^2\, \varepsilon^{-2}\, \ln^2 \varepsilon^{-1} \,+\, c(f)\, C\, M^2\, \varepsilon^2\, \ln^3 \varepsilon^{-1} \,+\, m\, C^2\, M^4\, \varepsilon^{-4}\, \ln^4 \varepsilon^{-1}\right)$$

with the big O-factor depending only on β.

For many kernels it is easy to check strong tractability of the conditional expectation problem independently of the transition densities. Indeed, define

$$K \,=\, \sup_{d=1,2,\dots}\; \sup_{x \in S_d} K_d(x,x). \tag{46}$$

Then since p_k is the density of the measure we have

$$0 \,\le\, \rho_{d,k} \,\le\, K \int_{S_d} p_k(x)\, dx = K.$$

Hence, $C \,\le\, 1 + K$. We summarize this in the corollary.

Corollary 1 *For uniformly bounded kernels , $K \,<\, \infty$, the conditional expectation problem for any Markov transition densities p_k is strongly tractable with strong exponent at most 2.*

Theorem 4 presents conditions on strong tractability of the conditional expectation and quasilinear contraction problems. Still, it is not entirely satisfactory due to the lack of constructive quadrature formulas. Also, it is not clear whether the strong exponent is really 2.

The problem of how to construct good quadrature formulas with an optimal exponent of ε^{-1} has been addressed in [33]. We briefly recall the property of this construction in a simplified case. We assume that the space F_d is the weighted tensor product of spaces of functions of one variable. That is, the domain S_d is now equal to D^d with D being a subset of $\mathbb{R}$. A typical example is $D = [0,1]$ which leads to the d-dimensional unit cube $S_d = [0,1]^d$. The reproducing kernel of F_d is now of the product form

$$K_d(x,t) \,=\, \prod_{k=1}^{d} \left(1 + \gamma_{d,k} K(x_k, t_k)\right), \tag{47}$$

where K is a reproducing kernel of the space of univariate functions. We assume that $K(\cdot, 0) = 0$. This assumption implies that the constant functions belong to F_d.

The weights $\gamma_{d,k}$ are nonnegative and moderate the behavior of functions for all variables. A small weight $\gamma_{d,k}$ means that the functions depend only slightly on the kth variable. The sum-exponent p_γ of the sequence $\gamma = \{\gamma_{d,k}\}$ is defined in [33]. Roughly speaking it is the largest positive number for which

$$\sup_d \sum_{k=1}^{d} \gamma_{d,k}^{p_\gamma} \; < \; \infty. \tag{48}$$

We assume that p_γ exists and $p_\gamma < 1$.

We also need to assume that integration is of the tensor product form, see (7) of [33]. This means that the transition densities are of the form

$$p_k(t) \; = \; \prod_{j=1}^{d} q_k(t_j) \tag{49}$$

for some one dimensional density $q_k : D \to \mathbb{R}$ from the space $L_2(D)$, and $k = 1, 2, \ldots, m$. The essence of (49) is that the d-dimensional density p_k is generated by the product of the one dimensional density q_k taken for the successive components of the vector t.

The *weighted tensor product* algorithm, shortly the WTP algorithm, is defined in [33] as approximation of general linear tensor product operators in the normalized sense. For the conditional expectation problem, the WTP algorithm is a quadrature formula of the form (27) which approximates $E_k(V)/\|E_k\|$.

Observe that we now have

$$\|E_k\|^2 \; = \; \|h_k\|^2 \; = \; \prod_{j=1}^{d} \left(1 + \gamma_{d,j} \int_{D^2} K(x,t) q_k(x) q_k(t) \, dt \, ds\right).$$

Observe that $\sup_d \sum_{j=1}^{\infty} \gamma_{d,kj} < \infty$ implies that all $\|E_k\|$ are of order 1. Hence, the normalized sense studied in [33] is equivalent to approximation of $E_k(V)$ which is needed for our purpose.

The WTP algorithm depends on a number of parameters. We may choose them in such a way that the WTP algorithm integrates the constant functions exactly. The error formula of the WTP algorithm has the following property. For any positive δ there exists a positive C_δ and there is the WTP algorithm which is a quadrature formula of the form (27) for which the error bound (28) is

$$\varepsilon_j(V) \;=\; C_\delta\, j^{-1/p^*} \min_{c\in\mathbb{R}} \|V-c\|_{F_d}, \qquad \text{with} \qquad p^* \;=\; \max\left(p+\delta, \frac{2p_\gamma}{1-p_\gamma}\right). \tag{50}$$

Here, p is the exponent of ε^{-1} for the one dimensional case, $d=1$. We stress that both C_δ and p^* do not depend on d. In particular, if

$$p_\gamma \;\le\; \frac{p}{p+2} \tag{51}$$

then

$$p^* \;=\; p+\delta.$$

In this case we can achieve the exponent p^* which is arbitrarily close to the one dimensional exponent p. Hence, we obtain the construction of quadrature formulas with the best possible exponent of ε^{-1}. This with Theorems 3 and 4 yield the following corollary.

Corollary 2 *Consider the spaces F_d with the reproducing kernel (47) and the weights $\gamma_{d,k}$ satisfying (51). Then the conditional expectation problem is strongly tractable with strong exponent p which is the exponent of ε^{-1} for the univariate case.*

Suppose that

$$C \;=\; \sup_{d,i=1,2,\dots} \frac{\min_{c\in R}\|V_i-c\|_{F_d}}{\|V^*\|_{F_d}}.$$

is finite. Then the quasilinear contraction problem is strongly tractable. The algorithm A defined by (29) with the WTP algorithm as $\hat{E}_{k,j}$ computes an ε-approximation and its cost satisfies

$$\mathrm{cost}(A) = O\left(c(u)\, m\, \alpha\, \varepsilon^{-p^*}\, \ln^{p^*}\varepsilon^{-1} \;+\; c(f)\,\alpha\,\varepsilon^{-p^*}\, \ln^{p^*+1}\varepsilon^{-1} \;+\; m\,\alpha^2\,\varepsilon^{-2p^*}\, \ln^{2p^*}\varepsilon^{-1}\right),$$

with $\alpha = CM^{p^*}$ where p^* can be arbitrarily close to the strong exponent of the conditional expectation problem, and the big O-factor depends only on β.

We now specify the results on strong tractability for the weighted Sobolev space $F_d = W^{1,1,\dots,1}([0,1]^d)$, see also [26, 33]. For this space we have $S_d=[0,1]^d$ and

$$K_d(x,t) \;=\; \prod_{i=1}^{d} \left(1+\gamma_{d,i}\min(x_i,t_i)\right).$$

The inner product of F_d is

$$\langle V, W\rangle \;=\; \sum_{u\,\subset\,\{1,2,\dots,d\}} \gamma_{d,u}^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial x_u} V(x_u,0) \frac{\partial^{|u|}}{\partial x_u} W(x_u,0)\, dx_u. \tag{52}$$

Here, $|u|$ is the cardinality of u. For the vector $x \in [0,1]^d$, we denote x_u as the vector from $[0,1]^{|u|}$ containing the components of x whose indices are in u, and $dx_u = \prod_{j\in u} dx_j$. By $(x_u, 0)$ we mean the vector x from $[0,1]^d$, with all components whose indices are not in u replaced by 0.

For $u = \emptyset$ we have $\gamma_{d,\emptyset} = 1$, and for $u \neq \emptyset$ we have $\gamma_{d,u} = \prod_{j\in u} \gamma_{d,j}$. If the weight $\gamma_{d,j}$ is zero then all $\gamma_{d,u} = 0$ with $j \in u$. In this case, we assume that the functions do not depend on the jth variable, and we have $0/0 = 0$ in the inner product formula.

Observe that the sum in the inner product has 2^d terms. For $\gamma_u \equiv 1$ we obtain the inner product of the classical Sobolev space. It is well known that the exponent p which appears in Corollary 2 is now 1, see [16].

The norms $\|V\|_{F_d}$ and $\|V\|$ may be quite different. Indeed, take the function $V(x) = (1-x_1)\dots(1-x_d)$. Then $\|V\| = 1$ and

$$\|V\|_{F_d} \;=\; \sum_u \gamma_u^{-1} \;=\; 1 + \sum_{j=1}^{d} \gamma_{d,j}^{-1} \prod_{k=j+1}^{d} \left(1 + \gamma_{d,k}^{-1}\right),$$

see formula (40) in [26]. Hence, for $\gamma_{d,k} = 1$ we have $\|V\|_{F_d} = 2^d$.

We first discuss strong tractability of the conditional expectation problem. Observe that

$$K_d(x,x) = \prod_{k=1}^{d} \left(1 + \gamma_{d,k} x_k\right).$$

The maximum of this function is attained for $x = [1,1,\dots,1]$ and K given by (46) is $K = \prod_{k=1}^{d}(1+\gamma_{d,k})$. This is finite iff $\sup_d \sum_{i=1}^{d} \gamma_{d,i} < \infty$.

Assume then that $\sup_d \sum_{i=1}^{d} \gamma_{d,i} < \infty$. In this case we have $p_\gamma \le 1$. Then the conditional expectation problem for any Markov densities is strongly tractable and the strong exponent is at most 2. With the additional assumption (51) that $p_\gamma \le 1/3$ we can apply Corollary 2 and the strong exponent is 1.

We add that it is known that the assumption $p_\gamma \le 1/3$ is *not* sharp for the Markov transition density $p_k \equiv 1$. It is proved in [11] that we can achieve the strong exponent is 1 assuming that $p_\gamma \le 1/2$. The proof is, however, not constructive.

We now turn to strong tractability of the quasilinear contraction problem. This holds under the addition assumption that $\sup_i \min_{c\in R} \|V_i - c\|_{F_d}/\|V^*\|_{F_d} < \infty$. We now discuss when this assumption holds. For simplicity, we consider the case when the transition densities are the same, $p_k(t|s) \equiv p(t)$ for some transition density p. As already mentioned in (17) and (18) we now have

$$\Gamma(V)(s) \;=\; \pi(s) \;+\; \beta \int_{S_d} V(t)p(t)\,dt$$

and the solution is

$$V^*(s) \;=\; \pi(s) \;+\; \frac{\beta}{1-\beta} \int_{S_d} \pi(t)p(t)\,dt, \qquad \forall\, s \in S_d.$$

Observe that the integrals $E_k(V)$ in (9) as well as quadrature formulas $\hat{E}_{k,j}$ in (27) do not now depend on k and s. Hence, $\hat{E}_{k,j}(V) = \hat{E}_j(V)$, and the iteration (29) takes now the form

$$V_i(s) \;=\; \pi(s) + \beta\, \hat{E}_{j_{i-1}}(V_{i-1}).$$

The functions V_i as well as the solution V^* differ from the function g only by constants. We also have $V_0 = 0$ and $V_1 = \pi$, where π is defined by (16). Therefore $V_i \in F_d$ for all i iff $\pi \in F_d$.

Assume then that $\pi \in F_d$. Then V^* also belongs to F_d and

$$\min_{c\in\mathbb{R}} \|V_i - c\|_{F_d} \;\equiv\; \min_{c\in\mathbb{R}} \|V_1 - c\|_{F_d} \;\le\; \|V^*\|_{F_d}.$$

This proves that Corollary 2 holds with $M = 1$.

We end this case of the transition densities by a simple remark that the algorithm A may be even further simplified by the use of the explicit form of the solution V^*. That is, we may take

$$A(s) \;=\; \pi(s) + \frac{\beta}{1-\beta}\, \hat{E}(\pi)$$

with an appropriate quadrature $\hat{E}$. Then we do not need to iterate and the bound of Corollary 2 holds without the logarithms of $\ln \varepsilon^{-1}$.

It is interesting to ask when the function π belongs to F_d. For simplicity assume that $m = 1$, take the quasilinear function $f(x) = x$ and the function $\pi = u_1$ as the Cobb-Douglas function,

$$\pi(x) \;=\; \prod_{k=1}^{d} (x_k + a_k)^{\alpha_k},$$

where $x = [x_1, x_2, \ldots, x_d]$ with nonnegative a_k and α_k such that $\sum_{k=1}^{d} \alpha_k = 1$.

Suppose we set $\gamma_{d,k} = \alpha_k$. It is shown in [33] that $\pi \in F_d$ iff $a := \min_j a_j > 0$. If so then

$$\|\pi\|_{F_d} \;\le\; b^2 + \frac{b^2}{a^2} \exp\left(1/a^2\right), \qquad \forall\, d,$$

where $b = \max_j a_j$. In particular, for $a_j \equiv 1$ we have $\|\pi\|_{F_d} \le 1 + \exp(1)$.

5 General Markov transition densities

In this section we consider Markov transition densities $p_k(t|s)$ which may depend on the second argument s. We relate strong tractability of the conditional expectation and quasilinear contraction problems to the *approximation* problem. By the approximation problem in a space F_d we mean approximation of elements V from F_d by using a finitely many function values of V. That is, $V(s)$ is approximating by the linear algorithm[2]

$$\hat{V}_j(s) \;=\; \sum_{i=1}^{j} b_{i,j}(s) V(t_{i,j}) \tag{53}$$

for some sample points $t_{i,j} \in S_d$ and some functions $b_{i,j}$ from the space $L_2(S_d)$. Let

$$e(\hat{V}_j) \;=\; \left\| V - \sum_{i=1}^{j} b_{i,j}(\cdot) V(t_{i,j}) \right\|_{L_2(S_d}$$

---

[2] It is known that more general algorithm such as nonlinear algorithms using adaptive choice of sample points are not better than linear algorithms considered in this section, see [30, 31].

be the error of the linear algorithm $\hat{V}_j$ for V. As for the problems studied in the previous sections, let $n_{\rm a}(\varepsilon,d)$ be the smallest integer n for which there exits $\hat{V}_n$ such that

$$e(\hat{V}_n) \ \le\ \varepsilon\,\|V\|_{F_d}, \qquad \forall\, V \in F_d.$$

We say that the approximation problem is *strongly tractable* in F_d iff there exist nonnegative C and p such that

$$n_{\rm a}(\varepsilon,d) \ \le\ C\,\varepsilon^{-p}, \qquad \forall\,\varepsilon\in(0,1),\ d=1,2,\dots\,. \tag{54}$$

The smallest (or infimum of) such p is called the *strong exponent* of approximation.

There are a number of papers where strong tractability of approximation in various classes of function is considered, see e.g., [35, 33], and a survey can be found in [31]. In particular, for some cases we know necessary and sufficient conditions under which strong tractability of approximation holds. This is sometimes achieved by assuming more general evaluations of V than function values but we do not pursue this point here.

We now relate strong tractability of approximation to strong tractability of the conditional expectation and quasilinear contraction problems.

Assume that (54) hold with $C=C_{\rm a}$ and $p=p_{\rm a}$. This means that for every d and j there exist sample points $t_{i,j}$ and functions $b_{i,j}$ such that the corresponding $\hat{V}_j$ satisfies

$$e(\hat{V}_j) \ \le\ C_{\rm a}^{1/p_{\rm a}}\,j^{-1/p_{\rm a}}\,\|V\|_{F_d}, \qquad \forall\, V\in F_d.$$

Knowing $\hat{V}_j$ we define the quadrature formula $\hat{E}_{k,j}$ as

$$\hat{E}_{k,j}(V)(s) \ =\ \int_{S_d}\hat{V}_j(t)\,p_k(t|s)\,dt \ =\ \sum_{i=1}^{j}\left(\int_{S_d} b_{i,j}(t)\,p_k(t|s)\,dt\right)\,V(t_{i,j}). \tag{55}$$

Hence, $\hat{E}_{k,j}$ is of the form (27) with

$$a_{i,j,k}(s) \ =\ \int_{S_d} b_{i,j}(t)\,p_k(t|s)\,dt \quad \text{and} \quad t_{i,j,k}\ =\ t_{i,j}.$$

In this case the sample points do not depend on k. The coefficients $a_{i,j,k}(s)$ can be precomputed for $s\in T_{\varepsilon,d}\cup\{s\}$ with the set $T_{\varepsilon,d}$ given by (33) with $t_{p,j_{i-1},k}=t_{p,j_{i-1}}$. Clearly,

$$E_k(V)(s)-\hat{E}_{k,j}(V)(s) \ =\ \int_{S_d}\Big(V(t)-\hat{V}_j(t)\Big)\,p_k(t|s)\,dt$$

and therefore

$$\left|E_k(V)(s) - \hat{E}_{k,j}(V)(s)\right| \,\le\, e(\hat{V}_j)\,\|p_k(\cdot|s)\|_{L_2(S_d)}.$$

Let

$$P \,=\, \sup_{k\in[1,m],\, s\in S_d} \|p_k(\cdot|s)\|_{L_2(S_d)}. \tag{56}$$

If P is finite then the upper bound $\varepsilon_j(V)$ of the quadrature error given by (28) is now given by

$$\varepsilon_j(V) \,=\, C_{\mathrm{a}}^{1/p_{\mathrm{a}}}\, j^{-1/p_{\mathrm{a}}}\, P\, \|V\|_{F_d}.$$

Hence, $\varepsilon_j(V) \le \varepsilon \|V\|_{F_d}$ if

$$j \,=\, C_{\mathrm{a}}\, P^{p_{\mathrm{a}}}\, \varepsilon^{-p_{\mathrm{a}}}.$$

This proves that the conditional expectation problem in F_d is strongly tractable with at most the same strong exponent as for approximation. This and Theorem 3 yield the following theorem.

Theorem 5 *If P given by (56) is finite then strong tractability of approximation in F_d implies strong tractability of the conditional expectation problem in F_d with at most the same strong exponent.*

If additionally M given by (40) is finite then the quasilinear contraction problem in F_d is strongly tractable with at most the same strong exponent of information and with the strong exponent of arithmetic operations at most twice as the strong exponent of approximation.

Furthermore, the algorithm A defined by (29) with $\hat{E}_{k,j}$ given by (55) computes an ε-approximation at cost

$$\mathrm{cost}(A) \,=\, O\left(c(u)\, m\, \alpha\, \varepsilon^{-p_{\mathrm{a}}}\, \ln^{p_{\mathrm{a}}} \varepsilon^{-1} \,+\, c(f)\, \alpha\, \varepsilon^{p_{\mathrm{a}}}\, \ln^{p_{\mathrm{a}}+1} \varepsilon^{-1} \,+\, m\, \alpha^2\, \varepsilon^{-2p_{\mathrm{a}}}\, \ln^{2p_{\mathrm{a}}} \varepsilon^{-1}\right)$$

with $\alpha = C_{\mathrm{a}}(MP)^{p_{\mathrm{a}}}$ and the big O-factor depending only on β.

We illustrate Theorem 5 for the space F_d which was already considered in Section 4. This is the Hilbert weighted tensor product space of functions defined on $S_d = D^d$ with the reproducing kernel K_d given by (47). We assume that

$$\int_{D^2} K(x,t)\, dt\, dx \,<\, \infty.$$

Then the operator $H(V)(s) = \int_D K(x,s)V(x)\,dx$ is compact and nonnegative definite. Consider its eigenpairs (λ_i, η_i), $H\eta_i = \lambda_i \eta_i$, with orthonormal η_i and ordered eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq 0.$$

Let p_λ be the sum-exponent of the sequence $\{\lambda_i\}$ of eigenvalues defined by (48). As in Section 4, p_γ is the sum-exponent of the sequence $\{\gamma_{d,k}\}$ of weights. It is proven in [33] that the necessary condition on strong tractability of approximation in F_d is that both p_λ and p_γ are finite, and then the strong exponent of approximation is at least $2\max(p_\lambda, p_\gamma)$.

Assume then that p_λ and p_γ are finite. In this case, the WTP algorithm is also effective for approximation, see [33]. Its construction is based on algorithms for the univariate case, $d = 1$. Let us assume that for $d = 1$ we know algorithms that use n function values with error proportional to n^{-1/p_1} for some positive p_1. It is known that $p_1 \geq 2p_\lambda$ and for some spaces we can achieve $p_1 = 2p_\lambda$. These univariate algorithms are used as building blocks for the WTP algorithm for arbitrary d. The WTP algorithm is of the form (53) with the functions $b_{i,j}$ which are the product of functions of one variable. That is, $b_{i,j}(s) = \prod_{k=1}^d b_{i,j,k}(s_k)$ for some functions $b_{i,j,k}$ from the space $L_2(D)$ and s_k is the kth component of the vector s. If we assume that

$$p_\gamma \leq \frac{p_1}{2 + 2p_1} \tag{57}$$

then the WTP algorithm computes an ε-approximation at cost $C_\delta \varepsilon^{p_1+\delta}$. Here, δ is positive and can be made arbitrarily small and C_δ is independent of d and may only depend on δ. This means that approximation is strongly tractable in F_d with strong exponent at most p_1. If $p_1 = 2p_\lambda$ then the strong exponent of approximation is exactly equal to $2p_\lambda$. This and Theorem 5 yields the following corollary.

Corollary 3 *Consider the spaces F_d with the reproducing kernel (47) and the weights $\gamma_{d,k}$ satisfying (57). Then approximation in F_d is strongly tractable with strong exponent at most p_1, where p_1 is the exponent of ε^{-1} for the univariate case.*

If P given by (56) is finite then the conditional expectation problem in F_d is strongly tractable with strong exponent at most p_1.

If additionally M given by (40) is finite then the quasilinear contraction problem in F_d is strongly tractable with strong exponent of information at most p_1 and strong exponent of arithmetic operations at most $2p_1$.

Furthermore, the algorithm A defined by (29) with the WTP algorithm to obtain $\hat{E}_{k,j}$ *by (55) computes an* ε*-approximation and its cost satisfies*

$$\mathrm{cost}(A) = O\left(c(u)\, m\, \alpha\, \varepsilon^{-p^*}\, \ln^{p^*} \varepsilon^{-1} \,+\, c(f)\, \alpha\, \varepsilon^{-p^*}\, \ln^{p^*+1} \varepsilon^{-1} \,+\, m\, \alpha^2\, \varepsilon^{-2p^*}\, \ln^{2p^*} \varepsilon^{-1}\right),$$

with $\alpha = C_\delta (MP)^{p^*}$ *where* $p^* = p_1 + \delta$*, and the big O-factor depends only on* β*.*

As in the previous section, we now specify the results for the weighted Sobolev space $F_d = W^{1,1,\dots,1}([0,1]^d)$. We now have $p_1 = 1$ and (57) means that $p_\gamma \le 1/4$.

Consider the case when the transition densities are the same $p_k(t|s) = p(t|s)$. Due to (16), the quasilinear contraction problem takes now the form

$$V(s) \;=\; \pi(s) + \beta \int_{[0,1]^d} V(t) p(t|s)\, dt. \tag{58}$$

The WTP algorithm generates the quadrature formulas $\hat{E}_{k,j} = \hat{E}_j$ which are now independent on k such that

$$\hat{E}_j(V)(s) \;=\; \sum_{i=1}^{j} a_{i,j}(s) V(t_{i,j}) \qquad \text{with} qquad a_{i,j}(s) \;=\; \int_{[0,1]^d} b_{i,j}(t) p(t|s)\, dt.$$

The condition $p_\gamma \le 1/4$ guarantees that for any $\delta \in (0,1)$ there exists a positive C such that we can choose the sample points $t_{i,j}$ and the functions $b_{i,j}$ for which

$$e_j(V) \;\le\; \min\left(C j^{-1+\delta}, 1-\beta\right)\, \|V\|_{L_2([0,1]^d)}, \tag{59}$$

see Theorem 5 of [33] applied to the problem of approximating functions V from F_d.

We want to check when V_i and V^* belong to F_d. Assume that $p(t|\cdot)$ belongs to F_d for all $t \in [0,1]$, and $\pi \in F_d$. Then $a_{i,j}$, $\hat{E}_j(V_i)$ as well as V_i belong to F_d for all i. The solution V^* also belongs to F_d since $p(t|\cdot) \in F_d$ for all $t \in [0,1]^d$ and

$$V^*(s) \;=\; \pi(s) + \beta \int_{[0,1]^d} V^*(t) p(t|s)\, dt$$

implies that all partial derivatives $\partial^{|u|} V^* / \partial x_u$ belong to $L_2([0,1]^{|u|})$.

We now estimate the ratios $\|V_i\|_{F_d}/\|V^*\|_{F_d}$. For any $u \in \{1,2,\dots,d\}$ by V^u we mean V if $u=\emptyset$ and $\partial^{|u|}V/\partial x_u$ otherwise. Similarly p^u denotes $\partial^{|u|}p(t|\cdot)/\partial x_u$. We have

$$\begin{aligned} V_i^u(s) &= \pi^u(s) + \beta\, \hat{E}^u_{j_{i-1}}(V_{i-1})(s), \\ (V^*)^u(s) &= \pi^u(s) + \beta \int_{[0,1]^d} V^*(t) p^u(t|s)\, dt. \end{aligned}$$

Recall that $E(V)(s) = \int_{[0,1]^d} V(t)p(t|s)\,dt$ is the integral of V with the weight p. Observe that the quadrature $\hat{E}^u_j$ is the usual quadrature applied to the integration problem $E^u(V)$. Clearly,

$$\|E^u(V) - \hat{E}^u_j(V)\|_{L_2([0,1]^d)} \le e_j(V)\ \|p^u\|_{L_2([0,1]^{2d})}\,.$$

Hence,

$$\|E(V) - \hat{E}_j(V)\|_{F_d} \le e_j(V)\, \|p\|_{F_d \times L_2([0,1]^d},$$

where

$$\|p\|_{F_d \times L_2([0,1]^d)} = \left(\int_{[0,1]^d} \|p(t|\cdot)\|^2_{F_d}\, dt \right)^{1/2}.$$

Since $V_i - V^* = \beta(\hat{E}_{j_{i-1}}(V_{i-1}) - E(V^*)) = \beta(\hat{E}_{j_{i-1}}(V_{i-1}) - E(V_{i-1}) + E(V_{i-1} - V^*))$ we have

$$\|V_i^u - (V^*)^u\|_{L_2([0,1]^d)} \le \beta\, \left(e_{j_{i-1}}(V_{i-1}) + \|V_{i-1} - V^*\|_{L_2([0,1]^d)} \right)\, \|p^u\|_{L_2([0,1]^d}. \qquad (60)$$

¿From (59) we have $e_j(V) \le (1-\beta)(\|V - V^*\|_{L_2([0,1]^d)} + \|V^*\|_{L_2([0,1]^d)})$. For $u = \emptyset$ we have $\|p^\emptyset\|_{L_2([0,1]^d} = 1$ and therefore

$$\|V_i - V^*\|_{L_2([0,1]^d)} \le \beta(2-\beta)\|V_{i-1} - V^*\|_{L_2([0,1]^d)} + \beta\|V^*\|_{L_2([0,1]^d)}.$$

This yields

$$\|V_i - V^*\|_{L_2([0,1]^d)} \le \frac{1-\beta+\beta^2}{1-\beta(2-\beta)}\, \|V^*\|_{L_2([0,1]^d)}.$$

Using (60) for all subsets u of $\{1,2,\dots,d\}$ we obtain

$$\|V_i - V^*\|_{F_d} \le \beta\, \left(e_{j_{i-1}}(V_{i-1}) + \|V_{i-1} - V^*\|_{L_2([0,1]^d)} \right)\, \|p\|_{F_d \times L_2([0,1]^d)}$$

and

$$\|V_i - V^*\|_{F_d} \le C_\beta\, \|V^*\|_{L_2([0,1]^d)}\, \|p\|_{F_d \times L_2([0,1]^d)}$$

with $C_\beta = (1 + (2-\beta)(1-\beta^2))/(1-\beta(2-\beta))$. This proves that

$$\sup_{i=1,2,\dots} \|V_i\|_{F_d}/\|V^*\|_{F_d} \le (1 + C_\beta)\, \|p\|_{F_d \times L_2([0,1]^d)} \frac{\|V^*\|_{L_2([0,1]^d)}}{\|V^*\|_{F_d}}.$$

It is known, see [33], that for any $V \in F_d$ we have

$$\|V\|_{L_2([0,1]^d)} \;\leq\; \prod_{k=1}^{d} \left(1 + \frac{4}{\pi^2}\gamma_{d,k}\right) \, \|V\|_{F_d}$$

Hence as long as $p_\gamma \leq 1$ then the ratio $\|V\|_{L_2([0,1]^d)}/\|V\|_{F_d}$ is of order 1, and $\|V_i\|_{F_d}/\|V^*\|_{F_d}$ is of order $\|p\|_{F_d \times L_2([0,1]^d)}$.

We summarize the results of this section. For the weighted Sobolev space F_d consider the set of transition densities

$$P_L \;=\; \left\{ p \,:\, \sup_{s\in[0,1]^d} \|p_k(\cdot|s)\|_{L_2([0,1]^d} \;\leq\; L \text{ and } \|p\|_{F_d\times L_2([0,1]^d)} \leq L \right\}$$

with a constant $L \geq 1$, as well as the set of functions

$$U \;=\; \{\, (u_1, u_2, \ldots, u_k) \;:\; \; f\,(u_1(\cdot), u_2(\cdot), \ldots, u_k(\cdot)) \in F_d \,\}\,.$$

Then Theorem 2 and the results of this section yield the following theorem.

Theorem 6 *If $p_\gamma \leq 1/4$ then the conditional expectation and quasilinear contraction problems with data from P and U are strongly tractable with the strong exponent 1. For any positive δ, the algorithm A defined by (29) with an appropriately chosen WTP algorithm computes an ε-approximation with*

$$\begin{aligned} \text{cost}(A) \;=\; O\Big(c(u)\, m\, L^{2(1+\delta)}\, \varepsilon^{-1-\delta}\, \ln^{1+\delta} \varepsilon^{-1} \quad &+ \quad c(f)\, m\, L^{2(1+\delta)}\, \varepsilon^{-1-\delta}\, \ln^{2+\delta} \varepsilon^{-1} \\ &+ \quad m\, L^{4(1+\delta)}\, \varepsilon^{-2(1+\delta)}\, \ln^{2(1+\delta)} \varepsilon^{-1} \Big) \end{aligned}$$

with the big O-factor depending only on β and δ.

References

[1] Anderson, E., L.P Hansen and T.J. Sargent Risk and robustness in general equilibrium manuscript, University of Chicago, 1998.

[2] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.*, 68, 337-404, 1950.

[3] R. Bellman *Dynamic Programming* Princeton University Press, Princeton, N.J. 1957.

[4] D. Blackwell Discounted Dynamic Programming *Annals of Mathematical Statistics* **36** 226–235, 1965.

[5] C. Chow and J. Tsitskilis, The complexity of dynamic programming, *Journal of Complexity* **5** 466–488, 1989.

[6] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, (Second edition), Academic Press, London, 1984.

[7] E. Denardo Contraction Mappings Underlying the Theory of Dynamic Programming *SIAM Review* **9** 165–177, 1967.

[8] L.G. Epstein and S. Zin Substitution, Risvk Aversion and the Temporal Behavior of Consumption and Asset Returns *Econometrica* **62-3** 937–969, 1989.

[9] W.H. Fleming and W. M. McEneaney Risk-sensitive control on an infinite time horizon *SIAM Journal of Control and Optimization* **33-6** 1881–1915, 1995.

[10] S. Heinrich Monte Carlo Complexity of Global Solution to Integral Equations *Journal of Complexity* **14-2** 151-175, 1998.

[11] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimensions, submitted for publication, 1998.

[12] Z. Huang, L. Khachiyan and K. Sikorski, Approximating fixed points of weakly contractive mappings, to appear in *J. Complexity*, 1999.

[13] R.E. Lucas, Jr. Asset prices in an exchange economy *Econometrica* **46-6** 1429–1446, 1978.

[14] D. McFadden "Econometric Models of Probabilistic Choice" in C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data* MIT Press, Cambridge, Massachusetts, 1981.

[15] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods* SIAM CBMS-NSF Monograph 63, Philadelphia, PA.

[16] E. Novak, *Deterministic and stochastic error bounds in numerical analysis*, Lecture Notes in Mathematics, Springer Verlag, Berlin, 1349, 1988.

[17] E. Novak and H. Woźniakowski, On the cost of uniform and nonuniform algorithms, 1997, to appear in *Th. Computer Science*.

[18] A. Papageorgiou and J.F. Traub, Beating Monte Carlo *Risk* 1996.

[19] A. Papageorgiou and J.F. Traub, Faster evaluation of multivariate integrals *Computations in Physics* 1997.

[20] S. Paskov and J. Traub, Faster valuation of financial derivatives, *Journal of Portfolio Management* **22** 113–120, 1995.

[21] W.H. Press, S.A. Teukolsky, W.T. Vettering, and B.P. Flannery *Numerical Recipes* Second Edition, Cambridge University Press, 1992.

[22] J.Rust, Stationary Equilibrium in a Market for Durable Assets *Econometrica* **53–4**, 783-805, 1985.

[23] J. Rust, Structural Estimation of Markov Decision Processes in R. Engle and D. McFadden (eds.) *Handbook of Econometrics* Volume 4, 3082–3139, North Holland, 1994.

[24] J. Rust, Using randomization to break the curse of dimensionality, *Econometrica*, **65-3**, 487–516 1997.

[25] J. Rust, A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations, manuscript, Yale University (available at `http://econwpa.wustl.edu/eprints/comp/papers/9704/9704001.abs`).

[26] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *J. Complexity*, 1-33, 14, 1998.

[27] K. Sikorski and H. Woźniakowski, Complexity of fixed points I, *J. Complexity*, 388-405, 3, 1987.

[28] G. Tauchen and R. Hussey, Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models *Econometrica* **59-2**, 371–396, 1991.

[29] S. Tezuka, *Uniform random numbers: theory and practice*, Kluwer Academic, Dordrecht, 1995.

[30] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.

[31] J. F. Traub, and A. G. Werschulz, *Complexity and Information*, Cambridge University Press, Cambridge, 1998.

[32] G. Wahba, Spline Models for Observational Data, SIAM-NSF Regional Conference Series in Appl. Math, 59, *SIAM*, Philadelphia.

[33] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor product algorithms for linear multivariate problems, submitted for publication, 1998.

[34] A. G. Werschulz, *The computational complexity of differential and integral equations*, Oxford University Press, New York, 1991.

[35] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity*, 96-128, 10, 1994.

[36] H. Woźniakowski, Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals, submitted for publication.

Author's Address:

John Rust, Department of Economics, Yale University, P.O. Box 208264, New Haven, CT 06520-8264,
email: `jrust@gemini.econ.yale.edu`,
http: `gemini.econ.yale.edu`,
Joseph F. Traub, Department of Computer Science, Columbia University,
New York, NY 10027, USA, email: `traub@cs.columbia.edu`,
Henryk Woźniakowski, Department of Computer Science, Columbia University,
New York, NY 10027, USA, and
Institute of Applied Mathematics and Mechanics, University of Warsaw,
ul. Banacha 2, 02-097 Warszawa, Poland, email: `henryk@cs.columbia.edu`