

# Is There a Curse of Dimensionality for Contraction Fixed Points in the Worst Case?\*

J. Rust, *Economics, Yale University*

J. F. Traub, *Computer Science, Columbia University*

H. Woźniakowski, *Computer Science, Columbia University and  
Institute of Applied Mathematics, University of Warsaw*

August 11, 2001

## Abstract

This paper analyzes the complexity of the *contraction fixed point problem*: compute an  $\varepsilon$ -approximation to the fixed point  $V^* = \Gamma(V^*)$  of a contraction mapping  $\Gamma$  that maps a Banach space  $B_d$  of continuous functions of  $d$  variables into itself. We focus on *quasi linear contractions* where  $\Gamma$  is a nonlinear functional of a finite number of conditional expectation operators. This class includes contractive Fredholm integral equations that arise in asset pricing applications and the contractive Bellman equation from dynamic programming. In the absence of further restrictions on the domain of  $\Gamma$ , the quasi linear fixed point problem is subject to the curse of dimensionality, i.e., in the worst case the minimal number of function evaluations and arithmetic operations required to compute an  $\varepsilon$ -approximation to a fixed point  $V^* \in B_d$  increases exponentially in  $d$ . We show that the curse of dimensionality disappears if the domain of  $\Gamma$  has additional special structure. We identify a particular type of special structure for which the problem is *strongly tractable* even in the worst case, i.e., the number of function evaluations and arithmetic operations needed to compute an  $\varepsilon$ -approximation of  $V^*$  is bounded by  $C\varepsilon^{-p}$  where  $C$  and  $p$  are constants independent of  $d$ . We present examples of economic problems that have this type of special structure including a class of rational expectations asset pricing problems for which the optimal exponent  $p = 1$  is nearly achieved.

---

\*This research was supported in part by the Alfred P. Sloan Foundation. J. F. Traub and H. Woźniakowski were supported in part by the NSF.

# 1 Introduction

This paper analyzes the complexity of the *contraction fixed point problem*: compute an  $\varepsilon$ -approximation to the fixed point  $V^* = \Gamma(V^*)$  of a contraction mapping  $\Gamma$  that maps a Banach space  $B_d$  into itself where  $B_d$  is a space of continuous functions defined on a compact  $S_d \subset \mathbf{R}^d$ . These problems arise frequently in economics, and include infinite horizon dynamic programming (DP) and asset pricing problems where  $V^*(s) = V^*(s_1, \dots, s_d)$  is interpreted as a *value function* that depends on a vector of  $d$  *state variables*  $(s_1, \dots, s_d)$  each of which can assume a continuum of possible values. Although the space  $B_d$  is infinite-dimensional we refer to  $d$  as the *dimension* of the fixed point problem. Although Banach's theorem guarantees that a unique fixed point exists, it is only possible in rare cases to obtain analytic expressions for  $V^*$  and it is usually quite difficult even to characterize general properties of  $V^*$  (e.g., monotonicity, concavity, etc.). For these reasons there has been increasing emphasis on the use of numerical methods to compute approximate contraction fixed points in virtually all fields in economics and finance, including microeconomic models of intertemporal decision making under uncertainty, Rust (1997), Judd (1998), econometrics, Rust (1994), macroeconomics, Stokey and Lucas (1989), Cooper (1999), growth theory, Kydland and Prescott (1982), and finance and asset pricing, Lucas (1978), Rust (1985), Duffie (1988).

In many applications we would like to be able to compute approximations to extremely high dimensional problems. In finance there is substantial interest in pricing index options that are functions of thousands of individual securities: calculation of theoretical values of these options requires solution of high dimensional optimal stopping problems where  $d$  is at least as large as the number of securities in the index, see Bolye, Broadie and Glasserman (1997). It is increasingly common to see very high dimensional DP in economic applications. For example, models of optimal pricing and inventory decisions of retail or wholesale companies can easily result in DP problems for which  $d$  can be many thousands. Hall and Rust (1999) model a steel wholesaler that carries over 2,000 individual steel products: each product is described by at least two continuous state variables, namely the current spot price and quantity on hand. Thus, to model the firm as a whole would potentially require a DP problem with at least  $d = 4,000$  continuous state variables. Similarly, Aguirregabiria (1999) develops a DP model of a Spanish retail supermarket chain that carries over 8,000 separate products: many U.S. retail outlets carry more than 75,000 products. Thus, it is increasingly important to find effective algorithms for approximating fixed points to high dimensional problems where  $d$  can be arbitrarily large.

Our analysis focuses on a subclass of *quasi linear contraction mappings* which can be

represented as nonlinear functionals of a finite number of conditional expectation operators:

$$\Gamma(V)(s) = f(\pi_1(s) + \beta E_1 V(s), \dots, \pi_m(s) + \beta E_m V(s)), \quad (1)$$

where  $f$  is a continuous function mapping  $R^m$  into  $R$ , the  $\pi_i$ ,  $i = 1, \dots, m$ , are fixed functions in  $B_d$ , and the  $E_i$ ,  $i = 1, \dots, m$ , are linear conditional expectation operators on  $B_d$ , i.e.,

$$E_i V(s) = \int_{S_d} V(t) p_i(t|s) dt, \quad (2)$$

where  $p_i : S_d \times S_d \rightarrow R_+$ , and  $p_i(\cdot|s)$  is a conditional probability density function on  $S_d$  for each  $s \in S_d$ . If  $f$  satisfies the quasi linearity property described in Section 2, then  $\Gamma$  is a contraction mapping on  $B_d$ . Many of the fixed point problems that appear in economic applications are quasi linear. This family includes Fredholm integral equations that arise in rational expectations models of asset pricing (when  $m = 1$  and  $f(x) = x$ ) and the Bellman equation for infinite horizon DP problems (when  $m > 1$  and  $f(x_1, \dots, x_m) = \max(x_1, \dots, x_m)$ ).

At least since Bellman (1957), it has been thought that contraction fixed point problems of the type we are considering are subject to an unavoidable “curse of dimensionality”. Subsequent developments in the theory of *computational complexity* of continuous problems, see Traub and Werschulz (1998) for a survey, have formalized the concept of curse of dimensionality and have succeeded in determining whether a wide variety of continuous mathematical problems are subject to this curse. The complexity,  $\text{comp}(\varepsilon, d)$ , of a  $d$ -dimensional mathematical problem is defined as the minimal cost of computing an  $\varepsilon$ -approximation in the worst case using deterministic algorithms.<sup>1</sup> If  $\text{comp}(\varepsilon, d)$  depends polynomially on  $\varepsilon^{-1}$  and  $d$ , we say the problem is *tractable*, otherwise it is *intractable*. If  $\text{comp}(\varepsilon, d)$  depends exponentially on  $d$ , we say the problem suffers the *curse of dimensionality*.

Chow and Tsitsiklis (1989) confirmed Bellman’s conjecture by proving that the problem of approximating the fixed point to Bellman’s equation  $V = \Gamma(V)$  is subject to the curse of dimensionality. More precisely, they showed that when domain of  $\Gamma$  is sufficiently large (e.g., the set  $L_d \subset B_d$  of all uniformly bounded Lipschitz continuous functions); then the problem of finding an  $\varepsilon$ -approximation to the fixed point  $V = \Gamma(V)$  is intractable. They considered a general class of Markovian decision problems with continuous state and action spaces. However their result also implies that the problem of approximating  $V^*$  for a subclass of MDPs with only a finite number  $m$  of possible actions known as *discrete decision processes* (DDPs) is also intractable, since its complexity is at least of order  $\varepsilon^{-2d}$ . However DDP problems are

---

<sup>1</sup>Complexity is also defined for average case and randomized settings. See Traub, Wasilkowski and Woźniakowski (1988).

mathematically equivalent to computing the fixed point of the Bellman operator which is just a special type of quasi linear contraction mapping. Thus, the intractability of the DDP problem implies that the quasi linear contraction problem (which includes the finite action MDP problem as a special case), must be intractable as well.

The objective of this paper is to determine circumstances under which it is possible to break intractability of the quasi linear contraction fixed point problem. As discussed in Traub and Werschulz (1998), there are two main ways this can be done:

- (a) by using randomized algorithms, such as Monte Carlo, or
- (b) by restricting attention to problems that have some sort of “special structure”.

Strategy (a) was used by Rust (1997) who showed that a random multigrid algorithm succeeds in breaking the curse of dimensionality for a subclass of DDP problems.<sup>2</sup> Since DDPs are a subclass of quasi linear contractions, Rust’s result suggests that it may be possible to use randomization to break the curse of dimensionality for a larger class of quasi linear contraction mappings. The main work involved in computing a fixed point in either of these problems is the evaluation of the multivariate integrals defining the conditional expectations  $E_i V$ ,  $i = 1, \dots, m$ . Although the problem of computing multivariate integrals for integrands that are in sufficiently large class such as  $L_d$  is intractable in the worst case deterministic setting, it is well known that multivariate integration is tractable when randomized algorithms are allowed. For example the classical Monte Carlo method can be used to approximate the integral  $\hat{E}_i V(s)$  as a sample average  $1/n \sum_{i=1}^n V(\tilde{s}_i)$  where  $\{\tilde{s}_1, \dots, \tilde{s}_n\}$  are *IID* draws from the conditional density  $p_i(t|s)$ . The Law of Large Numbers implies that  $\hat{E}_i V(s)$  converges to  $E_i V(s)$  at rate  $n^{-1/2}$  independent of the number of variables.<sup>3</sup> The cost of allowing randomization is that we must be content with the weaker stochastic assurance that the *expected* error of the approximate solution is less than  $\varepsilon$ .

Our paper is about strategy (b). There are two main advantages of this approach. First, unlike Monte Carlo methods, the algorithms we consider deliver a stronger deterministic

---

<sup>2</sup>It is possible to check that the results of Rust (1997) are also valid for the class of problems considered in this paper.

<sup>3</sup>Note that computer implementations of the Monte Carlo method use pseudo-random number generators that attempt to emulate truly *IID* draws from various distributions. The fact that the observed rate of convergence of these implementations of the Monte Carlo method is  $n^{-1/2}$  can be taken as evidence that pseudo-random number generators are able to successfully emulate the behavior of truly *IID* sequences. The problem of when the use of pseudo-random sample points does not change the error of Monte Carlo is addressed in Traub and Woźniakowski (1992).

worst case assurance regarding the approximation error. Second, we provide sufficient conditions under which there is a deterministic algorithm that uses  $n$  quadrature points to approximate the conditional expectation operators  $E_i V$  in equation (1) and results in an approximate fixed point  $\hat{V}_n$  that, regardless of the dimension  $d$ , converges to  $V^*$  at rate close to  $n^{-1}$ .

We now describe the type of special structure we exploit and how it enables us to break the curse of dimensionality. Our objective is to approximate  $V^*$  at an arbitrary point  $s^* \in S_d$ . Note that  $V^*$  is an implicit function of the objects  $f$ ,  $\{p_i\}$  and  $\{\pi_i\}$ , where  $f$  is the function defining the quasilinear contraction operator in equation (1), the  $\{\pi_i\}$  are the payoff functions in (1), and the  $\{p_i\}$  are the Markov transition densities defining the conditional expectation operators in (2). We assume that a feasible algorithm for computing an approximate solution  $\hat{V}$  to  $V^*$  can only depend on the values of these objects at a finite number of points  $\{s_1, \dots, s_n\}$  in the domain  $S_d$ . Our worst case analysis treats the objects  $f$  and  $\{p_i\}$  as fixed, but allows  $\{\pi_i\}$  to be elements of a set of functions  $F_d$ . For any algorithm (which includes a specific choice for  $n$  and the sample points  $\{s_1, \dots, s_n\}$ ) we can determine the worst case error, i.e., the largest possible value of  $|V^*(s^*) - \hat{V}(s^*)|$  for  $\{\pi_i\}$  in the set  $F_d$ . The complexity  $\text{comp}(\varepsilon, d)$  is the minimum cost of computing an  $\varepsilon$ -approximation, where the minimum is taken over all feasible algorithms. Obviously, the complexity depends on the set of functions  $F_d$ . When  $F_d = B_d$ , the (non-compact) set of all continuous, uniformly bounded functions, the complexity is infinite, i.e., it is not possible to compute an  $\varepsilon$ -approximation for sufficiently small  $\varepsilon$  using deterministic algorithms. When  $F_d = L_d$ , the compact set of uniformly bounded Lipschitz-continuous functions, the complexity is finite, but as noted above, there is a curse of dimensionality —  $\text{comp}(\varepsilon, d)$  is an exponential function of  $d$ . We will consider the case where  $F_d = F_{d,\gamma}$ , a compact subset of uniformly bounded functions in a weighted Sobolev space that will be described in more detail later. We will define a specific *weighted norm*  $\|\cdot\|_{F_{d,\gamma}}$  on the functions in  $F_{d,\gamma}$  that depends on a parameter  $\gamma = (\gamma_1, \dots, \gamma_d)$  that has the same dimension as the of arguments  $d$  of the functions in  $F_{d,\gamma}$ . The  $i^{\text{th}}$  component  $\gamma_i$  is a weight that determines, for any function  $\pi \in F_{d,\gamma}$ , how sensitive the norm  $\|\pi\|_{F_{d,\gamma}}$  is to variations in  $s_i$ , the  $i$  argument of  $\pi$ . If  $\gamma_i$  is small the norm will be very large if the function  $\pi$  is highly variable as a function of  $s_i$ . Thus, in order for  $\pi$  to have a small weighted norm when  $\gamma_i$  is small, it must be the case that  $\pi$  must be nearly “flat” as a function of  $s_i$ . We show that the quasi linear contraction fixed point problem becomes tractable in the worst case setting if  $\sum_{d=1}^{\infty} \gamma_d < \infty$ .

Tractability is established by showing that  $\text{comp}(\varepsilon, d)$  takes the form

$$\text{comp}(\varepsilon, d) = c(d)n_1(\varepsilon, d) + n_2(\varepsilon, d), \quad (3)$$

where  $n_1(\varepsilon, d)$  is the minimal number of function evaluations,  $n_2(\varepsilon, d)$  is the minimal number

of arithmetic operations required to compute an  $\varepsilon$ -approximation to a  $d$ -dimensional problem in the worst case, and  $c(d)$  represents the cost of a single function evaluation of  $\pi$ . Clearly, the problem is tractable if  $n_1$ ,  $n_2$  and  $c$  all increase polynomially in  $\varepsilon^{-1}$  and  $d$ . We restrict attention to problems where  $c(d)$  is bounded by a polynomial in  $d$ , since if this does not hold, the problem is automatically intractable. In the quasilinear contract problem, we have  $n_2(\varepsilon, d) = O(n_1(\varepsilon, d)^2)$ , so the key to our argument is to show that  $n_1$  is a polynomial in  $\varepsilon^{-1}$  and  $d$ . We provide sufficient conditions under which the problem is *strongly tractable*, i.e., where

$$n_1(\varepsilon, d) \leq C\varepsilon^{-p} \quad (4)$$

for constants  $C$  and  $p$  which are *independent* of  $d$  and  $\varepsilon$ . The minimal (or the infima of) values of  $p$  for which these inequalities hold is called the *strong exponent*, see Woźniakowski (1994) for a precise definition. We provide sufficient conditions under which the strong exponent is equal to  $p = 1$  which implies that the optimal (uni-dimensional) convergence rate of  $n^{-1}$  is nearly achieved regardless of the problem dimension  $d$ .

We verify that these rates of convergence can be attained in the context of a specific rational expectations asset pricing example. In this case the quasi linear contraction problem reduces to a Fredholm integral equation for  $V^*$ , where  $V^*(s)$  represents the value of the asset given information  $s$ , where  $s$  is a vector in a  $d$ -dimensional state space  $S_d$ . The dimension  $d$  could be quite large, since it not only includes a vector of variables that represent the state of the firm (i.e., its sales, costs, debt, wage costs, market share, and so forth), but it also includes a large number of market-level variables summarizing the status of the firm's chief competitors, and a large number of economy-level variables that affect the overall level of the stock market (e.g., interest rates, inflation rates, unemployment rates, and so forth). Suppose that the asset pays a per share dividend of  $\pi(s)$  in state  $s$ , and that the state evolves according to a Markov transition density  $p$  and there is a constant risk free interest rate  $r > 0$ . Then the “fundamental value” of the stock is given by

$$V^*(s) = \pi(s) + \frac{1}{(1+r)} \int_{S_d} V^*(t)p(t|s)dt. \quad (5)$$

In Section 5 we show that the problem of approximating  $V^*$  is strongly tractable if  $\pi$  satisfies some mild smoothness conditions and  $p$  is a truncated normal density for which certain parameters related to the eigenvalues of its correlation matrix tend to infinity at a sufficiently fast rate. We prove that this problem is strongly tractable with strong exponent equal to 1. That is, the number of function evaluations  $n(\varepsilon, d)$  necessary to compute an  $\varepsilon$ -approximation to  $V^*$  at a point satisfies  $n_1(\varepsilon, d) \leq C_\delta \varepsilon^{-1-\delta}$  for all  $\delta > 0$ . This is substantially faster than

the randomized contraction operator approach considered by Rust (1997). In that algorithm  $n_1(\varepsilon, d)$  must be of order  $\varepsilon^{-2}$  in order to achieve an *expected* error of  $\varepsilon$ . Thus, even though  $V^*$  is a non-trivial function of all  $d$  state variables, the number of sample points necessary to find an  $\varepsilon$ -approximation is independent of  $d$ . The problem complexity,  $\text{comp}(\varepsilon, d)$ , only depends on  $d$  via the cost  $c(d)$  of each function evaluation.<sup>4</sup> This gives us a double-win:

- convergence is faster than Monte Carlo,
- with a worst case deterministic guarantee.

Although we do not do so here, our results on the tractability of pointwise approximation of  $V^*$  can be extended to enable us to prove the tractability of uniform approximation of  $V^*$ . Consider a function  $\hat{V}_k$  constructed from a linear combination of  $k$  “basis functions”  $\{\rho_{i,k}\}$  whose coefficients are values  $\hat{V}$  evaluated at an appropriate set of sample points  $\{s_1, \dots, s_k\}$  in  $S_d$

$$\hat{V}_k(s) = \sum_{i=1}^k \hat{V}(s_i) \rho_{i,k}(s). \quad (6)$$

$\hat{V}_k$  will be a *uniform  $\varepsilon$ -approximation* to  $V^*$  provided  $\|V^* - \hat{V}_k\|_1 \leq \varepsilon \|V^*\|_2$  for some norms  $\|\cdot\|_1$  and  $\|\cdot\|_2$ . The value of  $k$  for which this is true obviously depends on  $\varepsilon$  and  $d$ . In order for the uniform approximation problem to be tractable we must guarantee that  $k$  depends only polynomially on  $1/\varepsilon$  and  $d$ , and that the time required to evaluate each of the basis functions depends only polynomially on  $d$ . This can be achieved for the same conditions under which the pointwise approximation problem is tractable, i.e., when the functions  $\hat{V}$  and  $V^*$  have a particular type of special structure reflected by small norms in the weighted Sobolev space  $F_{d,\gamma}$ . Thus, it is sufficient to restrict attention to proving that the problem of pointwise approximation of  $V^*$  is tractable.

Section 2 introduces the class of quasi linear contraction mappings, with examples of how these mappings arise in economic applications. Section 3 reviews recent empirical and theoretical results on Quasi Monte Carlo (QMC). This is relevant to our results since we show how a particular QMC algorithm that is based on an approximation algorithm known

---

<sup>4</sup>Werschulz (1991) and Heinrich (1998) analyzed the worst case complexity of Fredholm integral equations for smooth functions and showed that these problems are subject to a curse of dimensionality. However their results do not necessarily imply that the quasi linear contraction problem suffers the curse of dimensionality since the class of Fredholm integral equations they consider include problems with more general kernels. Our analysis only covers a subset of *contractive* Fredholm integral equations where the kernel takes the form  $\beta p(t|s)$  where  $\beta \in (0, 1)$  and  $p(\cdot|s)$  is a conditional probability density function with special structure.

as the *weighted tensor product algorithm* (WTP) enables us to exploit the special structure of integrands in the Sobolev space  $F_{d,\gamma}$  and break the curse of dimensionality. In Section 4 we present a class of iterative algorithms for computing approximate fixed points to quasi linear contractions. Section 5 outlines the structure of our argument in the special *IID* case where the transition densities  $p_i(t|s)$  do not depend on  $s$ . Section 6 applies these algorithms to the problem of approximating fixed points to an important subclass of quasi linear mappings, (contractive) Fredholm integral equations. These problems arise in asset pricing applications such as discussed above, and also in the *policy iteration* method for solving DDPs (i.e., Markov decision processes with a finite number of possible actions). Section 7 verifies that under mild restrictions, the multidimensional rational expectations asset pricing problem is tractable with strong exponent equal to  $p = 1$ . Section 8 provides a short conclusion which discusses some of the limitations of our results and offers suggestions for future research in this area. The appendix contains the more technical definitions and proofs.

## 2 The Quasilinear Contraction Problem

**Definition 1.** A function  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  is quasi linear if:

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

**Examples:**

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

$$f(x) = \sum_{k=1}^m c_k x_k, \tag{7}$$

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

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

$$f(x) = \max_{k=1,2,\dots,m} x_k. \tag{8}$$



3. *Smoothed Max function.* Let  $f_\sigma$  be given by

$$f_\sigma(x) = \sigma \ln \left( \sum_{k=1}^m \exp(x_k/\sigma) \right) \quad (9)$$

for a nonnegative  $\sigma$ .

The direct use of (9) is not recommended in computation due to overflow problems for small  $\sigma$ . Note, however, that

$$f_\sigma(x) = \max_{k=1,2,\dots,m} x_k + \sigma \ln \left( \sum_{i=1}^m \left( \exp \left( (x_i - \max_{1 \leq k \leq m} x_k) / \sigma \right) \right) \right).$$

This formula can be used for computation of  $f_\sigma(x)$  without problems of overflow since  $x_i - \max_{k=1,2,\dots,m} x_k \leq 0$  and we always compute the exponential for a non-positive argument. Observe also that

$$0 \leq f_\sigma(x) - \max_{k=1,2,\dots,m} x_k \leq \sigma \ln m, \quad (10)$$

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

4. *“Social Surplus functions”.* Let  $f_\sigma$  be given by

$$f_\sigma(x) = \int_{\mathbf{R}^m} \max_{k=1,2,\dots,m} [x_k + \sigma \xi_k] q(\xi_1, \dots, \xi_m) d\xi_1 \cdots d\xi_m, \quad (11)$$

where  $q(\xi_1, \dots, \xi_m)$  is a probability density over  $\mathbf{R}^m$  which has finite absolute first moments. This class of functions was introduced in McFadden (1981). The adjective “Social Surplus” reflects the interpretation of  $f_\sigma(x)$  as the expected utility of a population of agents indexed by  $\xi = (\xi_1, \dots, \xi_m)$  who face  $m$  possible choices, where the utility of choice  $k$  is  $x_k + \sigma \xi_k$ . The smoothed max function (9) is a special case of (11) when  $q$  is the product of  $m$  appropriately standardized univariate Type III extreme value distributions, i.e.,

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

where  $\gamma \simeq .577216\dots$  is Euler’s constant.

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

$$\pi_k : S_d \rightarrow \mathbf{R} \quad \text{for} \quad k = 1, 2, \dots, m \quad (12)$$

be elements of  $B_d$ , and let

$$p_k : S_d \times S_d \rightarrow \mathbf{R}_+ \quad \text{for} \quad k = 1, 2, \dots, m \quad (13)$$

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

$$E_k V(s) = \int_{S_d} V(t) p_k(t|s) dt. \quad (14)$$

**Definition 2.** A mapping  $\Gamma : B_d \rightarrow B_d$  is **quasi linear** if  $\Gamma$  is given by:

$$\Gamma(V)(s) = f(\pi_1(s) + \beta E_1 V(s), \dots, \pi_m(s) + \beta E_m V(s)), \quad (15)$$

where  $f : R^m \rightarrow R$  is a quasi linear function, and  $\beta \in (0, 1)$ .

**Examples:**

1. *Fredholm Integral Equations*

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

$$\Gamma(V)(s) = \pi(s) + \beta EV(s), \quad \forall s \in S_d. \quad (16)$$

Thus  $\Gamma$  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, Lucas (1978) or Tauchen and Hussey (1991). Much is known about the complexity of computing approximate solutions to Fredholm integral equations with general kernels, see Werschulz (1991) and Heinrich (1998) for surveys of deterministic and stochastic complexity bounds for this problem. In the case of general kernels, the results in Werschulz (1991) show that the problem is intractable in the worst case using deterministic algorithms. By exploiting the special structure of the kernel for the class of Fredholm integral equations combined with additional special structure on the functions  $\pi_i$  and  $p_i$  (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  $f(x) = \max[x_1, \dots, x_m]$  and obtain

$$\Gamma(V)(s) = \max[\pi_1(s) + \beta E_1 V(s), \dots, \pi_m(s) + \beta E_m V(s)].$$

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., Blackwell (1965) and Denardo (1967)).

We stress that this function  $f$  corresponds to the case of finitely many choices of actions, and the case of a continuous choice of actions is not addressed. But even this restricted case poses difficulties for our analysis since our results depend on the ability to exploit smoothness properties of the function  $\Gamma(V)$ . Even if the functions  $\pi_k$  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. One possible solution to the problem is to approximate  $V^*$  via Howard's (1960) *policy iteration* algorithm. The algorithm consists of alternating policy improvement and policy valuation steps and is globally convergent. Each policy valuation step is equivalent to the solution of a Fredholm integral equation. We provide sufficient conditions for the tractability of approximating the fixed point to a class of Fredholm integral equations in Section 6. Assuming that the number of policy iterations required to converge grows only polynomially in  $d$ , it is possible to use an approximate policy iteration algorithm to approximate the fixed point  $V^*$  to Bellman's equation. This requires showing that the approximate policy iteration algorithm will converge provided the value function at each approximate policy valuation step is approximated sufficiently accurately. Another possible solution is to smooth the Bellman operator as discussed below.

## 3. Smoothed Bellman Operators.

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

$$\Gamma_\sigma(V)(s) = \sigma \ln \left[ \sum_{k=1}^m \exp \left\{ \frac{1}{\sigma} [\pi_k(s) + \beta E_k V(s)] \right\} \right]. \quad (17)$$

The function  $\Gamma(V)$  is as smooth as  $\pi_k$  and  $p_k$ . From (10) we conclude

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

and therefore the Bellman operator is a uniform limit of smoothed Bellman operators:  $\Gamma = \lim_{\sigma \rightarrow 0} \Gamma_\sigma$ . This justifies the name of  $\Gamma_\sigma$  as the smoothed Bellman operator. Fixed

point problems with smoothed Bellman operators arise in econometric applications, see e.g., Rust (1994).

#### 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  $f_\sigma$  given in equation (11),

$$\Gamma_\sigma(V)(s) = \int_{\mathbf{R}^m} \max_{k=1, \dots, m} [\pi_k(s) + \beta E_k V(s) + \sigma \xi_k] q(\xi_1, \dots, \xi_m) d\xi_1 \cdots d\xi_m. \quad (18)$$

It is obvious from (18) that  $\Gamma = \lim_{\sigma \rightarrow 0} \Gamma_\sigma$ . While our analysis of the complexity of the quasi linear 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 quasi linear function in (18). Thus, we will assume that the function  $f$  in equation (15) can be evaluated exactly, such as in the case of the smoothed Bellman operators (neglecting 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 the approximation error in the quasi linear function  $f$  in addition to the approximation error in  $E_k$ .

**Theorem 1** *Let  $\Gamma$  be a quasi linear mapping given in Definition 2. Then  $\Gamma$  is a contraction, and the equation*

$$V = \Gamma(V)$$

*has a unique solution  $V^*$  in the ball  $U_R = \{V \in B_d : \|V\| \leq R\}$  with  $R \geq (1 - \beta)^{-1} \|f(\pi_1(\cdot), \dots, \pi_m(\cdot))\|$ , and*

$$\frac{\|f(\pi_1(\cdot), \dots, \pi_m(\cdot))\|}{1 + \beta} \leq \|V^*\| \leq \frac{\|f(\pi_1(\cdot), \dots, \pi_m(\cdot))\|}{1 - \beta}.$$

*Proof:* . The quasi linearity property of  $f$  implies that  $\Gamma$  satisfies Blackwell's (1965) sufficient conditions for a contraction mapping. However it is actually easier to verify that  $\Gamma$  is a contraction via a direct calculation. We have  $\|E_k(V)\| \leq \|V\|$  and

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

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

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

Since  $\beta < 1$ ,  $\Gamma$  is a contraction mapping.

By Banach's theorem, the unique solution of  $V = \Gamma(V)$  in  $U_R$  exists if  $\Gamma(U_R) \subset U_R$ . The last inclusion holds since

$$|\Gamma(V)(s)| \leq f(\pi_1(s) + \beta\|V\|, \dots, \pi_m(s) + \beta\|V\|) = f(\pi_1(s), \dots, \pi_m(s)) + \beta\|V\|.$$

Hence, for  $V \in U_R$  we have

$$\|\Gamma(V)\| \leq \|f(\pi_1(\cdot), \dots, \pi_m(\cdot))\| + \beta R \leq R$$

due to the condition on  $R$ . The bounds on the norms of  $V^*$  easily follow from the pointwise estimates

$$-\beta\|V^*\| + f(\pi_1(s), \dots, \pi_m(s)) \leq V^*(s) \leq \beta\|V^*\| + f(\pi_1(s), \dots, \pi_m(s)).$$

□

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

$$\Gamma(V)(s) = \pi(s) + \beta \int_{S_d} V(t)p(t|s) dt, \quad \text{with } \pi(s) = f(\pi_1(s), \dots, \pi_m(s)). \quad (19)$$

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

$$\Gamma(V)(s) = \pi(s) + \beta \int_{S_d} V(t)p(t) dt. \quad (20)$$

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

$$V^*(s) = \pi(s) + \frac{\beta}{1 - \beta} \int_{S_d} \pi(t)p(t) dt. \quad (21)$$

Hence, computation of  $V^*$  reduces to the computation of a single multivariate integral in this case. This indicates that quasi linear contractions are computationally at least as hard as multivariate integration.

**Definition 3.** *The quasi linear contraction problem is defined as the problem of computing an  $\varepsilon$ -approximation to  $V^*(s^*)$  at a given point  $s^* \in S_d$  where  $V^*$  is the unique solution to the contraction fixed point problem*

$$V = \Gamma(V), \quad (22)$$

and  $\Gamma$  is a quasi linear contraction mapping satisfying Definition 2.

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

$$|V^*(s^*) - \hat{V}(s^*)| \leq \varepsilon \max(\|V^*\|, |||V^*|||). \quad (23)$$

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 quasi linear contraction problem 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  $\varepsilon$ -approximation is easier.

### 3 Strong Tractability of Multivariate Integration

As we will see later multivariate integration is the key “subproblem” of the quasi linear contraction problem and the potential source of a curse of dimensionality. In this section we review what is currently known about the complexity of multivariate integration, summarizing the circumstances under which the curse of dimensionality arises and when it can be “broken”. As already noted in the introduction, there are two main ways that this can be done: a) via randomized methods such as Monte Carlo, or b) by exploiting special structure of particular classes of integrands.

Consider the problem of computing the integral  $I(f)$  given by

$$I(f) = \int_{S_d} f(x)dx, \quad (24)$$

where for concreteness we assume that  $S_d$  is the  $d$ -dimensional cube  $S_d = [0, 1]^d$ . It has long been known that multivariate integration is subject to a curse of dimensionality in the worst case deterministic setting when the integrands are allowed to be members of a sufficiently general normed space of functions of  $d$  variables,  $F_d$ . Let the norm of the space  $F_d$  be denoted by  $\|\cdot\|_{F_d}$ . The curse of dimensionality can be explained in terms of the *error bound*  $e_n(F_d)$  given by

$$e_n(F_d) = \inf_{I_n} \sup_{f \in F_d, \|f\|_{F_d} \leq 1} |I(f) - I_n(f)|, \quad (25)$$

where  $I_n(f)$  represents some *algorithm* for computing an approximate integral using  $n$  evaluations of the integrand  $f$ .

We now present an example of  $F_d$  for which the curse of dimensionality is present. The space  $F_d = C^r([0, 1]^d)$  is the set of functions defined on  $S_d$  which are  $r$ -times Frechet continuously differentiable. That is, we assume that for any multi-index  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_d]$  with nonnegative integers  $\alpha_i$ , such that  $|\alpha| := \alpha_1 + \dots + \alpha_d \leq r$ , we have that

$$D^\alpha f = \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} x_1 \partial^{\alpha_2} x_2 \dots \partial^{\alpha_d} x_d} f$$

is continuous. The norm of  $f$  in  $C^r([0, 1]^d)$  is defined as the maximal value of derivatives of  $f$  up to order  $r$ ,

$$\|f\|_{C^r([0,1]^d)} = \max_{\alpha, |\alpha| \leq r} \max_{x \in S_d} |D^\alpha f(x)|.$$

Bakhvalov (1959) showed that for the space  $F_d$  we have  $e_n(F_d) = \Theta(n^{-r/d})$  and the complexity of the integration problem is given by

$$\text{comp}(\varepsilon, d) = \Theta \left( c(d) \left( \frac{1}{\varepsilon} \right)^{d/r} \right), \quad (26)$$

where  $c(d)$  is the cost of a single evaluation of  $f$  and the notation  $g = \Theta(h)$  means that the function  $g$  is asymptotically proportional to the function  $h$  as  $\varepsilon$  tends to zero, i.e.,  $g = O(h)$  and  $h = O(g)$ .<sup>5</sup>

Bakhvalov's result implies that for fixed  $r$ , the complexity depends exponentially on  $d$ , so multivariate integration is subject to an unavoidable curse of dimensionality using deterministic algorithms. We are interested in the question of whether there are smaller spaces of functions for which the curse of dimensionality disappears. For sufficiently small spaces, such as the class of all  $d$ -variate polynomials of order at most  $k$ , there are analytic formulas for the integral and the complexity of integration is finite even for  $\varepsilon = 0$ . In fact, the complexity is proportional to the number of polynomial terms which is of order  $d^k$ . Obviously, we would like to find non-trivial spaces  $F_d$  which are practically important, are as large as possible, and for which the curse of dimensionality for multivariate integration is not present. We believe that an example of such spaces is the case of *weighted Sobolev* spaces  $F_{d,\gamma}$  which will be discussed below. For these spaces the curse of dimensionality is *not* present, and even simple algorithms, such as QMC discussed below, will allow good error bounds which go to zero as a low degree polynomial in  $n^{-1}$  independently of  $d$ .

---

<sup>5</sup>Bakhvalov's original article is in Russian, however version of his theorem and proof can be also found in books written in English, see e.g., Novak (1988), Traub and Werschulz (1998) or Traub, Wasilkowski and Woźniakowski (1988).

Consider the class of Quasi Monte Carlo (algorithms). These are integration algorithms for which the numerical integral  $I_n(f)$  can be written as a simple weighted sum of the form

$$I_n(f) = \sum_{i=1}^n w_{i,n} f(s_{i,n}), \quad (27)$$

where  $\{w_{i,n}\}$  are  $n$  quadrature weights and  $\{s_{i,n}\}$  are  $n$  sample points in  $S_d$ . In the classical *Monte Carlo* algorithm (MC) and in *Quasi Monte Carlo* algorithms (QMC),  $w_{i,n} = 1/n$ . The difference between QMC and MC is that in QMC the sample points are deterministically chosen by a formula while in MC the sample points are random, independent and identically distributed (*IID*) draws from the uniform distribution on  $[0, 1]^d$ . It has long been known that the classical Monte Carlo algorithm succeeds in breaking the curse of dimensionality provided  $F_d$  is a set for which the associated standard deviation  $\sigma(f) = (I(f^2) - I^2(f))^{1/2}$  is polynomially bounded in  $d$ . This follows from the fact that

$$\sup_{f \in F_d} E^{1/2} \left\{ \left| I(f) - n^{-1} \sum_{i=1}^n f(\tilde{t}_i) \right|^2 \right\} = \frac{\sup_{f \in F_d} \sigma(f)}{\sqrt{n}}, \quad (28)$$

where the expectation is taken with respect to *IID* sample points  $\{\tilde{t}_i\}$ .

QMC algorithms differ from the classical Monte Carlo algorithm by attempting to *deterministically* choose a set of  $n$  sample points  $\{t_i\}$  which are as close to being uniform as possible. By *uniform* we mean that the fraction of points lying within any rectangular subregion (with sides parallel to the coordinate axes and containing zero) of the  $d$ -dimensional unit cube is as close as possible to the volume (Lebesgue measure) of that subregion. The *discrepancy* of a given set of  $n$  points  $\{t_1, \dots, t_n\}$  is a measure of their deviation from uniformity. The discrepancy can be measured in various ways, the most commonly used are  $L_2$ -discrepancy and  $L_\infty$ -discrepancy. The formal definition of the  $L_\infty$ -discrepancy  $D_\infty$  is given by

$$D_\infty(t_1, \dots, t_n) = \sup_{B \in \mathcal{B}} |\lambda_n(B) - \lambda(B)| \quad (29)$$

where  $\mathcal{B}$  is the set of all closed subintervals of  $[0, 1]^d$  (i.e., sets of the form  $B = \prod_{i=1}^d [0, b_i]$ ,  $b_i \in (0, 1]$ ),  $\lambda(B) = b_1 b_2 \cdots b_d$  is the Lebesgue measures of the set  $B$ , and  $\lambda_n(B)$  is the *empirical measure* of  $B$ :

$$\lambda_n(B) = \frac{1}{n} \sum_{i=1}^n 1_B(t_i), \quad (30)$$

where  $1_B$  is the indicator function  $B$ . That is,  $1_B(t) = 1$  if  $t \in B$ , and  $1_B(t) = 0$  if  $t \notin B$ . If instead of taking the maximum difference between  $\lambda_n(B)$  and  $\lambda(B)$  over all  $B \in \mathcal{B}$  we



compute the distance using the  $L_2$  or  $L_p$  norms we obtain the  $L_2$  and  $L_p$  discrepancies,  $D_2$  and  $D_p$ ,  $p \geq 1$ , respectively. We can do this by noting that there is a one to one mapping between the set  $B$  and the point  $b \in [0, 1]^d$  representing the vector of endpoints of  $B = \prod_{i=1}^d [0, b_i]$ . Then we can define the  $L_p$  discrepancy  $D_p$  as

$$D_p(t_1, \dots, t_n) = \left[ \int_{[0,1]^d} |\lambda_n(x) - \lambda(x)|^p dx \right]^{1/p} \quad (31)$$

As usual, the  $L_2$ -discrepancy is a special case when  $p = 2$  and the  $L_\infty$ -discrepancy is the limit of  $L_p$ -discrepancies as  $p \rightarrow \infty$ .

The *Koksma-Hlawka inequality* allows us to bound the error of the QMC algorithm by the discrepancy of the sample points  $\{t_i\}$ . It is usually stated in terms of the  $L_\infty$  discrepancy as

$$|I(f) - I_n(f)| \leq V(f)D_\infty(t_1, \dots, t_n), \quad (32)$$

where  $V(f)$  denotes the *variation* of the function  $f$  (for definition see Niederreiter, 1992). However in this paper, it is more convenient to work with a version of the Koksma-Hlawka inequality that can be stated in terms of the  $L_2$ -discrepancy:

$$|I(f) - I_n(f)| \leq \|f\|_{F_{d,1}} D_2(t_1, \dots, t_n), \quad (33)$$

where the norm  $\|f\|_{F_{d,1}}$  is used instead of the variation  $V(f)$ .<sup>6</sup> More precisely, this bound applies to  $f$  in the Sobolev space  $F_{d,1}$  of functions that are once differentiable with respect to each variable. This is a special case of the weighted Sobolev space  $F_{d,\gamma}$  introduced in the appendix with  $\gamma_{d,i} \equiv 1$ . The norm in the space  $F_{d,1}$  is denoted by  $\|\cdot\|_{F_{d,1}}$ . Here we only mention that for  $d = 1$  we have

$$\|f\|_{F_{1,1}}^2 = f(0)^2 + \int_0^1 f'(t)^2 dt, \quad (34)$$

whereas for  $d \geq 2$  we have

$$\|f\|_{F_{d,1}}^2 = \sum_{u \subset \{1,2,\dots,d\}} \int_{[0,1]^{|u|}} f^{(u)}(x_u, 0)^2 dx_u,$$

where for  $u = \{u_1, u_2, \dots, u_k\}$  with  $k = |u|$  the cardinality of the subset  $u$ ,  $x_u = (x_{u_1}, \dots, x_{u_k})$  and  $f^{(u)} = \partial^{|u|}/(\partial x_{u_1} \cdots \partial x_{u_k})$  the partial derivative with respect to  $x_{u_j}$ . Finally,  $(x_u, 0)$  is a

---

<sup>6</sup>The  $L_2$ -discrepancy is also related to the average case error of QMC for the class of continuous functions equipped with the classical Wiener sheet measure, see Woźniakowski (1991).

vector with the  $j$ th component equal to  $x_j$  if  $j$  is in  $u$ , and 0 otherwise.<sup>7</sup>

The Koksma-Hlawka inequality makes it clear why we are interested in QMC algorithms based on sample points that have low discrepancy. Formally, a sequence  $\{s_{i,n}\}$ ,  $i = 1, \dots, n$ ,  $n = 1, 2, \dots$ , is said to have *low discrepancy* if  $D_p$  satisfies

$$D_p(s_{1,n}, \dots, s_{n,n}) \leq a(d) \frac{\log^{d-1}(n)}{n} \quad (35)$$

for some positive  $a(d)$ , and  $p = 2$  or  $p = \infty$ .

Low discrepancy sequences (LDS) have been extensively studied in the last 40 years. Examples of LDS are Halton, Sobol, Faure, generalized Faure, and Niederreiter sequences, see Niederreiter (1992), Tezuka (1995), and Drmota and Tichy (1997). However QMC methods based on LDS have been thought to be inappropriate for high dimensional integration, since for large  $d$   $\log(n)^{d-1}/n$  may be substantially larger than  $1/\sqrt{n}$  unless  $n$  is huge. Therefore, up until recently it was widely believed that QMC should not be used for, say,  $d \geq 12$ . The only way to do very high dimensional integration seemed to be via classical Monte Carlo.

Then, in the mid-nineties, computer experiments on financial applications with  $d = 360$  showed that QMC beat MC by factors of 10 to 1000; see Paskov and Traub (1995), Papa-georgiou and Traub (1996), and Paskov (1997). Generalized Faure points seemed especially effective in these problems. Many numerical experiments showed that the convergence rate in these problems was roughly  $n^{-1}$  independently of  $d$ . A survey of computer experiments on financial instruments can be found in Traub and Werschulz (1998). In many economic and financial applications limited computing capacity and the need for results in “real time” limits the size of  $n$  to a few hundred or thousand, yet as we discussed in the introduction, the problem dimension  $d$  may be in the hundreds or thousands. The  $(\log n)^{d-1}$  factor in the error bound suggests that MC should beat QMC, yet precisely the opposite happened in the computer experiments.

The existing theory of QMC algorithms was unable to explain these experimental findings. The challenge was to develop a theory that could explain why QMC converges as  $n^{-1}$

---

<sup>7</sup>We stress that the space  $F_{d,1}$  is one of many examples of Sobolev spaces. Even assuming the same smoothness of functions one can define a different norm for the Sobolev space. For instance, one can take for  $d = 1$ , instead of (34),

$$\|f\|^2 = \int_0^1 f(t)^2 dt + \int_0^1 f'(t)^2 dt,$$

and use the tensor product norm for general  $d$ . This probably corresponds to the most popular Sobolev space which is often use in the study of differential equations. The results for multivariate integration in this latter Sobolev space (with weights) are basically the same as for the space  $F_{d,\gamma}$  as recently shown in Sloan and Woźniakowski (2000).

independently of  $d$  in these financial applications. This suggests the possibility that the Koksma-Hlawka inequality is too conservative, and if the type of problems that have been tested have some form of special structure it might be possible to derive a tighter bound on the integration error.

What special structure might these problems have? In many economic and financial problems we are interested in computing expected discounted values of future flows of dividends or utilities. Due to the effect of the additive separability and the discounting of payoffs or utilities, variables representing payoffs in more distant points in the future are less important than variables representing near-term payoffs. That is, the integrands are *non-isotropic*. Could this special structure be used to vanquish the curse of dimensionality with a worst case guarantee?

Sloan and Woźniakowski (SW,1998) formalized a particular type of special structure of functions that includes functions that can be represented as discounted sums and quantified how much it can help. They analyzed the error of integration for functions in the Sobolev space  $F_{d,\gamma}$  with a particular weighted norm,  $\|f\|_{d,\gamma}$ , see the appendix. The symbol  $\gamma$  refers to a sequence of weights  $\{\gamma_{d,i}\}$  where  $\gamma_{d,i}$  moderates the behavior of the functions of  $d$  variables with respect to the  $i$ th variable. The weight  $\gamma_{d,i}$  enters the norm inversely so that if an element of  $f$  has weighted norm, say, at most 1 then small  $\gamma_{d,i}$  means that the function  $f$  is almost “flat” with respect to the  $i$ th variable. If we reorder the arguments of  $f(x_1, \dots, x_d)$  to have non-increasing weights  $\gamma_{d,i}$ , then this is equivalent to assuming that arguments with successively higher indices have monotonically declining effects on the values of  $f$ . One can show that various parametric families of functions commonly used in economics, e.g., Cobb-Douglas production functions, belong to the class  $F_{d,\gamma}$ .

SW showed that there exist QMC algorithms for which the curse of dimensionality is *not* present in the worst case under certain conditions on  $\{\gamma_{d,i}\}$ . In order to summarize their key results, which are directly relevant for our analysis in this paper, we first recall the notions of tractability and strong tractability for multivariate integration (see Woźniakowski (1994) for a more in-depth discussion). Tractability can be defined in terms of the error bounds  $e_n(F_d)$ . For  $n = 0$  we do not sample the functions, and we set  $e_0(F_d) = \|I\|$  as the initial error. Suppose we want to reduce the initial error by a factor  $\varepsilon \in (0, 1)$ . Let  $n = n(\varepsilon, d)$  be the minimal  $n$  for which

$$e_n(F_d) \leq \varepsilon e_0(F_d).$$

We say that integration is *tractable* in  $F_d$  iff  $n(\varepsilon, d)$  can be bounded by a polynomial in  $d$  and  $\varepsilon^{-1}$ , and *strongly tractable* iff  $n(\varepsilon, d)$  can be bounded by a polynomial only in  $\varepsilon^{-1}$ . Otherwise, we say it is *intractable*.

For simplicity, we present results by assuming that  $\gamma_{d,i} = \gamma_i$  for the weighted Sobolev space  $F_{d,\gamma}$ . Then integration is strongly tractable in  $F_{d,\gamma}$  iff

$$\sum_{i=1}^{\infty} \gamma_i < \infty, \quad (36)$$

and integration is tractable in  $F_{d,\gamma}$  iff

$$\limsup_{d \rightarrow \infty} \frac{\sum_{i=1}^d \gamma_i}{\log d} < \infty. \quad (37)$$

SW proved this result for QMC algorithms, and Novak and Woźniakowski (1999b) proved it for arbitrary algorithms. Furthermore, if we consider the classical unweighted Sobolev space  $F_{d,\gamma}$  with  $\gamma_i \equiv 1$  then the complexity of integration depends exponentially on  $d$  and the curse of dimensionality is present.

If (36) holds then the number,  $n(\varepsilon, d)$ , of function evaluations plus arithmetic operations needed to compute an  $\varepsilon$ -approximation in the worst case setting satisfies

$$n(\varepsilon, d) \leq C \varepsilon^{-p} \quad (38)$$

where both  $C$  and  $p$  are independent of  $d$  and  $\varepsilon$ , 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, and the rate of convergence is at least as fast as the classical Monte Carlo algorithm. It is also known, due to Hickernell and Woźniakowski (1999), that

$$\sum_{i=1}^{\infty} \gamma_i^{1/2} < \infty \quad \text{implies} \quad p^* = 1, \quad (39)$$

where  $p^*$  is the infimum of  $p$  satisfying (38). Hence, with a more restrictive condition on  $\gamma_i$  we have the same rate of convergence as for the one-dimensional case.

The proofs of these results are non-constructive and they do not specify for which algorithms the bound (38) holds. If condition (39) on the weights is replaced by a stronger condition,

$$\sum_{i=1}^{\infty} \gamma_i^{1/3} < \infty, \quad (40)$$

then an algorithm satisfying (38) with  $p$  almost 1 has been constructed. This is the *weighted tensor product* (WTP) algorithm of Wasilkowski and Woźniakowski (1999). The WTP algorithm is defined for arbitrary multivariate linear problems and its definition can be found in

the appendix. We will use the WTP algorithm to approximate the conditional expectation operators  $E_i$  entering the quasi linear mapping  $\Gamma$ . We provide sufficient conditions under which the conditional expectation operators  $E_i$  map  $F_{d,\gamma}$  into itself and the conditional expectation problem is strongly tractable. Sections 4 and 5 will use this result and the WTP algorithm as the basis for an algorithm for solving the quasi linear contraction fixed point problem.

## 4 Algorithms

In this section we consider various algorithms for solving the quasi linear contraction problem. We begin by assuming for a moment that we can evaluate  $\Gamma(W)(s^*)$  exactly for a given function  $W$ . Then we can solve (22) by the simple iteration

$$V_i(s^*) = \Gamma(V_{i-1})(s^*), \quad i = 1, 2, \dots, \quad (41)$$

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

$$\|V_i - V^*\| \leq \beta^i \|V^*\|, \quad \forall i.$$

Hence  $V_i$  converges to the fixed point  $V^*$ . Let

$$\varepsilon_1 = \varepsilon \frac{\max(\|V^*\|, \|V^*\|)}{\|V^*\|}. \quad (42)$$

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

$$n = \left\lceil \frac{\ln 1/\varepsilon_1}{\ln 1/\beta} \right\rceil. \quad (43)$$

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  $\varepsilon_1 \geq \varepsilon$  and we may bound  $n$  by replacing  $\varepsilon_1$  by  $\varepsilon$ ,  $n \leq \lceil \ln(1/\varepsilon) / \ln(1/\beta) \rceil$ . The latter bound is constructive.

If  $\beta$  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  $\beta$ . These factors are of order 1 if  $\beta$  is not too close to 1. The case of  $\beta$  close to 1 is also of interest although it is not studied in this paper. For  $\beta$  close to 1, the iteration (41) as well as all its modifications studied in this paper can

be significantly improved for moderate values of  $d$  and  $\varepsilon^{-1}$ , and the number of steps can be proportional to  $\ln 1/(1 - \beta)$  as shown in Sikorski and Woźniakowski (1987), and Huang, Khachiyan and Sikorski (1999).

Thus, as long as  $\Gamma(V_n)(s^*)$  can be computed exactly, the quasi linear 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 weights  $p_k(\cdot|s^*)$ . This can be done, in general, only approximately.

Assume then that instead of (41) we can compute the perturbed sequence

$$V_i(s^*) = \Gamma(V_{i-1})(s^*) + \delta_{i-1}(s^*), \quad \text{with } |\delta_{i-1}(s^*)| \leq \delta_{i-1} \|V_{i-1}\|, \quad (44)$$

for some nonnegative  $\delta_{i-1}$ . We will see later that  $\delta_{i-1}$  corresponds to the quadrature error and can be made sufficiently small by taking sufficiently many sample points in the quadrature formula. For  $i = 1$ , we have  $V_0 = 0$  and there is no error in integration. Hence,  $\delta_0 = 0$ .

It is natural to ask how small  $\delta_{i-1}$  should be to preserve the global convergence property of the sequence (41). In what follows, we assume that

$$\beta + \delta_{i-1} \leq \bar{\beta} < 1, \quad \forall i = 1, 2, \dots \quad (45)$$

We have

$$\begin{aligned} \|V_i - V^*\| &\leq \beta \|V_{i-1} - V^*\| + \delta_{i-1} \|V_{i-1}\| \leq (\beta + \delta_{i-1}) \|V_{i-1} - V^*\| + \delta_{i-1} \|V^*\| \\ &\leq \bar{\beta} \|V_{i-1} - V^*\| + \delta_{i-1} \|V^*\|. \end{aligned}$$

This yields

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

If we set

$$\bar{\beta}^n \leq \varepsilon_1/2 \quad \text{which holds for} \quad n = \left\lceil \frac{\ln 2/\varepsilon_1}{\ln 1/\bar{\beta}} \right\rceil,$$

and

$$\bar{\beta}^{n-1-i} \delta_i \leq \frac{\varepsilon_1}{2n}, \quad i = 0, 1, \dots, n-1,$$

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

$$\|V_n - V^*\| \leq \varepsilon_1 \|V^*\| = \varepsilon \max(\|V^*\|, \|V^*\|).$$

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

$$\delta_i \leq \frac{\bar{\beta}^i \ln 1/\bar{\beta}}{\ln 2/\varepsilon_1 + \ln 1/\bar{\beta}} \leq \frac{\varepsilon_1}{2n\bar{\beta}^n} \bar{\beta}^{i+1}. \quad (46)$$

Hence,  $\delta_i$  may mildly depend on  $\varepsilon_1$ , 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 (41) we are ready to replace multivariate integrals in  $\Gamma(V)$  by quadrature formulas. We approximate the conditional expectation  $E_k$  given by (14) by quadrature formulas

$$\hat{E}_{k,j}(V)(s) = \sum_{i=1}^j a_{i,j,k}(s) V(t_{i,j,k}) \quad s \in S_d, \quad (47)$$

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,

$$\left| E_k(V)(s) - \hat{E}_{k,j}(V)(s) \right| \leq \varepsilon_j(V), \quad \forall s \in S_d, \quad k = 1, 2, \dots, m. \quad (48)$$

We are ready to modify the algorithm (41) 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 (41) is now formally given by

$$V_i(s^*) = f\left(\pi_1(s^*) + \beta \hat{E}_{1,j_{i-1}}(V_{i-1})(s^*), \dots, \pi_m(s^*) + \beta \hat{E}_{m,j_{i-1}}(V_{i-1})(s^*)\right), \quad \forall i = 1, 2, \dots \quad (49)$$

We are ready to prove

**Theorem 2** . *Assume that the quadrature errors satisfy*

$$\varepsilon_{j_i}(V_i) \leq \min \left( \frac{1-\beta}{2}, \frac{\left(\frac{1+\beta}{2}\right)^i \ln \frac{2}{1+\beta}}{\beta \left(\ln \frac{2}{\varepsilon_1} + \ln \frac{2}{1+\beta}\right)} \right) \|V_i\| \quad (50)$$

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 (49) is an  $\varepsilon$ -approximation to  $V^*(s^*)$ .

*Proof:* It is easy to check that (49) is of the form (44) with

$$\delta_{i-1} = \beta \frac{\varepsilon^{j_{i-1}}(V_{i-1})}{\|V_{i-1}\|}.$$

Let  $\bar{\beta} = (1 + \beta)/2$ . From (50) we conclude that (45) as well as (89) hold and therefore  $A(s^*)$  is an  $\varepsilon$ -approximation.  $\square$

We now discuss the cost of the algorithm  $A(s^*) = V_n(s^*)$  given by (49) and Theorem 2 with

$$n = \left\lceil \frac{\ln 2/\varepsilon_1}{\ln 2/(1 + \beta)} \right\rceil. \quad (51)$$

We assume that we can compute the values of the functions  $\pi_k$ ,  $k = 1, 2, \dots, m$ , and  $f$  at any point as well as that we can perform arithmetic operations. Let the cost of one evaluation of a function  $\pi_k$  be  $c(\pi_k)$  and let the cost of one evaluation of the function  $f$  be  $c(f)$ . Observe that  $\pi_k$  is a function of  $d$  variables and therefore the cost  $c(\pi_k)$  may depend on  $d$ . Let  $c(d) = \max_{k=1, \dots, m} c(\pi_k)$ . 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 as  $\beta a_{i,j_i,k}(s)$  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  $\varepsilon$  we do not include the cost of generating these precomputed numbers. This is a typical assumption in complexity analysis, see Traub, Wasilkowski and Woźniakowski (1988) and Novak and Woźniakowski (1999a) where this point is fully discussed. We also stress that precomputing significantly simplifies when we use the WTP algorithm, see the appendix.

We now explain in detail how  $A(s^*)$  can be computed. For  $i = 1, 2, \dots, n$ , and  $s \in S_d$  denote

$$V_i(s) = f\left(\pi_1(s) + \sum_{p=1}^{j_{i-1}} \beta a_{p,j_{i-1},1}(s) V_{i-1}(t_{p,j_{i-1},1}), \dots, \pi_m(s) + \sum_{p=1}^{j_{i-1}} \beta a_{p,j_{i-1},m}(s) V_{i-1}(t_{p,j_{i-1},m})\right). \quad (52)$$

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

$$T_{\varepsilon,d} = \left\{ t_{p,j_{i-1},k} : i \in [2, n], p \in [1, j_{i-1}], k \in [1, m] \right\} \quad (53)$$



denote all integration sample points used in (52), and let  $|T_{\varepsilon,d}|$  denote the cardinality of the set  $T_{\varepsilon,d}$ . Clearly,

$$|T_{\varepsilon,d}| \leq m (j_1 + j_2 + \cdots + j_{n-1}). \quad (54)$$

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

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

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

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

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  $V_1, V_2, \dots, V_{n-1}$  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 compute  $\pi_k(x)$  for all  $x \in T_{\varepsilon,d} \cup \{s^*\}$  and  $k \in [1, m]$  at cost at most  $m(1 + |T_{\varepsilon,d}|)c(d)$ . To compute  $V_i(x)$  for all  $x \in T_{\varepsilon,d}$ , we perform  $2j_{i-1}m$  arithmetic operations and one evaluation of the function  $f$  at cost  $(2j_{i-1}m + c(f))|T_{\varepsilon,d}|$ . Finally we compute  $V_n(s^*)$  at cost  $2j_{n-1}m + c(f)$ . Then  $\text{cost}(A)$  of computing  $A(s^*)$  is

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

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 (47). More precisely, the cost of  $A$  depends for which indices  $j$  we can guarantee that the integration error  $\varepsilon_j(V)$  satisfies (50). Note that (50) holds if we set

$$\varepsilon_{j_i}(V_i) = C_\beta \left( \frac{1 + \beta}{2} \right)^i \frac{\|V_i\|}{\ln \varepsilon_1^{-1}} \quad (55)$$

for some  $C_\beta$  depending only on  $\beta$ . Let

$$N(\varepsilon, d) = j_1 + j_2 + \cdots + j_{n-1} \quad (56)$$

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

$$\text{cost}(A) = O \left( N(\varepsilon, d) \left( m c(d) + (\ln \varepsilon^{-1}) c(f) \right) + N^2(\varepsilon, d) m \right) \quad (57)$$

with the big  $O$ -factor depending only on  $\beta$ .

The essence of (57) 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  $c(d)$  and  $c(f)$  and quadratic in  $N(\varepsilon, d)$  in terms of the cost of arithmetic operations. Typically, the cost  $c(d)$  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 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  $\varepsilon^{-1}$  independent of  $d$ .

Suppose we want to compute an  $\varepsilon$ -approximation to  $E_k(V)(s)$  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 (47) such that

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

We say that the conditional expectation problem is *strongly tractable*<sup>8</sup> in  $F_d$  if there exist nonnegative  $C$  and  $p$  such that

$$n(\varepsilon, d) \leq C \varepsilon^{-p}, \quad \forall \varepsilon \in (0, 1), d = 1, 2, \dots \quad (58)$$

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

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

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

Hence,  $p$  is the exponent of  $\varepsilon^{-1}$  which tells us how many evaluations of  $\pi_i$  and  $f$  are needed, and  $p_1$  is the exponent of  $\varepsilon^{-1}$  which tells how many arithmetic operations are needed to solve the quasi linear 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*

---

<sup>8</sup>It is also reasonable to study the case of *tractability* in which we permit polynomial dependence on  $d$ . For simplicity the focus of our attention in this paper is on strong tractability.

of the quasi linear contraction problem.<sup>9</sup>

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

$$j_i = C \left( C_\beta \frac{\|V_i\|}{\|V_i\|_{F_d}} \left( \frac{1+\beta}{2} \right)^i \frac{1}{\ln \varepsilon_1^{-1}} \right)^{-p} \quad (59)$$

with  $C$  and  $p$  given by (58), and  $C_\beta$  by (55). Then there exists  $\hat{E}_{k,j_i}$  of the form (47) such that

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

Hence (55) is satisfied. We now estimate  $N(\varepsilon, d)$  for  $j_i$  given by (59). Then

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

Using the definition of  $\varepsilon_1$ , see (42), we have

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

with the big  $O$ -factor depending only on  $\beta$ . Note that  $\|V^*\|$  is of order  $\|V_i\|$ . Hence  $\|V^*\|/\|V_i\|$  can be dropped 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

**Theorem 3** . *Algorithm A computes an  $\varepsilon$ -approximation to the solution of the quasi linear contraction problem at cost*

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

---

<sup>9</sup>For many *linear* problems, we can guarantee that  $p_1 = p$ . This is due to the fact that there exists a linear optimal error algorithm that requires the same order of arithmetic operations as the number of information evaluations, see Traub and Werschulz (1998) for a survey. In our case, we have a nonlinear problem for which it may happen that  $p$  and  $p_1$  are different. In fact, it may be a tradeoff between the cost of information and arithmetic operations. That is, it may happen that with the minimal number of information evaluation we must perform significantly more arithmetic operations, whereas the use of more information evaluations may allow to reduce the number of arithmetic operations. There are examples of such nonlinear problems, see Novak and Woźniakowski (1999).

where  $N(\varepsilon, d)$  is given by (56). Suppose that

$$M = \sup_{d,i=1,2,\dots} \frac{\|V_i\|_{F_d}}{\max(\|V^*\|, \|V^*\|_{F_d})} \quad (60)$$

is finite. Then strong tractability of the conditional expectation problem in  $F_d$  with (58) implies strong tractability of the quasi linear contraction problem and the cost of the algorithm  $A$  satisfies

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

with the big  $O$ -factor depending only on  $\beta$ .

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

Theorem 3 relates strong tractability of the conditional expectation problem and the quasi linear contraction problem as long as we know that  $V_i$  and  $V$  are in the space  $F_d$  and the bound  $M$  of (60) 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  $F_d = B_d$ . As noted in the introduction, if the functions in  $B_d$  are only continuous, the problem is not only intractable, it is insoluble for sufficiently small  $\varepsilon$  since the integration error  $e_n(F_d)$  do not converge to zero as  $n \rightarrow \infty$  when we have the freedom to choose a worst case integrand from the class of all continuous functions, see e.g., Traub, Wasilkowski and Woźniakowski (1988). If we want to use the space  $B_d$ , one possible solution is to follow Rust (1997) and switch to the randomized setting in which the bound (48) is understood as the expected error with respect to randomized sample points. 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 quasi linear 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.

## 5 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 quasi linear contraction problem are strongly tractable.

For the densities  $p_k$  independent of  $s$ , the assumption on the quality of the quadrature rules (48) 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

$$\left| \int_{S_d} V(t) p_k(t) dt - \sum_{i=1}^n a_{i,j,k} V(t_{i,j,k}) \right| \leq \varepsilon_j(V) \quad k = 1, 2, \dots, m. \quad (61)$$

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  $K_d : S_d \times S_d \rightarrow R$ . A reproducing kernel Hilbert space is defined as one where the value of a function  $V$  at a point  $t$  is given by the inner product of  $V$  and the kernel  $K_d(\cdot, t)$  evaluated at  $t$ . More precisely, 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 the basic theory of such spaces the reader is referred to Aronszajn (1950) and Wahba (1990) as well as to Section 9.1 of the appendix.

For  $k = 1, 2, \dots, m$ , define

$$h_k(x) = \int_{S_d} K_d(x, t) p_k(t) dt, \quad x \in S_d. \quad (62)$$

Assuming that  $h_k \in F_d$  we have from (14) and (47)

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

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 (61) 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  $a_{i,j,k} = 1/j$ . 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

$$\rho_{d,k} = \int_{S_d} K_d(x, x) p_k(x) dx - \int_{S_d^2} K_d(x, t) p_k(x) p_k(t) dx dt. \quad (63)$$

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 (47) satisfies

$$\left| E_k(V) - \hat{E}_{k,j}(V) \right| \leq \|V\|_{F_d} \frac{\sqrt{\rho_{d,k}}}{\sqrt{j}}. \quad (64)$$

We stress that the proof of (64) 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 (64) proves that (58) holds with  $p = 2$  and

$$C = \sup_{k \in [1, m], d=1, 2, \dots} (1 + \rho_{d,k}) \quad (65)$$

as long as  $C$  is finite. We thus have

**Theorem 4 .** *Suppose that  $C$  given by (65) is finite. Then the conditional expectation problem is strongly tractable with strong exponent at most 2. Suppose additionally that  $M$  given by (60) is finite. Then the quasi linear 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 (49) with (non-constructive)  $\hat{E}_{k,j}$  computes an  $\varepsilon$ -approximation at cost*

$$\text{cost}(A) = O \left( c(d) m \alpha \varepsilon^{-2} \ln^2 \varepsilon^{-1} + c(f) \alpha \varepsilon^{-2} \ln^3 \varepsilon^{-1} + m \alpha^2 \varepsilon^{-4} \ln^4 \varepsilon^{-1} \right)$$

with  $\alpha = CM^2$ , and the big  $O$ -factor depending only on  $\beta$ .

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

$$K = \sup_{d=1, 2, \dots} \sup_{x \in S_d} K_d(x, x). \quad (66)$$

Then since  $p_k$  is the density of the measure we have

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

Hence,  $C \leq 1 + K$ . We summarize this in

**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 gives sufficient conditions for strong tractability of the conditional expectation and quasi linear contraction problems. Still, it is not entirely satisfactory due to the lack of constructive quadrature formulas. Also, the strong exponent might be less than 2.

The problem of how to construct good quadrature formulas with an optimal exponent of  $\varepsilon^{-1}$  has been addressed in Wasilkowski and Woźniakowski (1999). We briefly summarize 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  $\mathbf{R}$ . 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

$$K_d(x, t) = \prod_{k=1}^d (1 + \gamma_{d,k} K(x_k, t_k)), \quad (67)$$

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  $k$ th variable. The sum-exponent  $p_\gamma$  of the sequence  $\gamma = \{\gamma_{d,k}\}$  is defined in Wasilkowski and Woźniakowski (1999). Roughly speaking it is the largest positive number for which

$$\sup_d \sum_{k=1}^d \gamma_{d,k}^{p_\gamma} < \infty. \quad (68)$$

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

We also need to assume that integration is of the tensor product form, see (7) of Wasilkowski and Woźniakowski (1999). This means that the transition densities are of the form

$$p_k(t) = \prod_{j=1}^d q_k(t_j) \quad (69)$$

for some one dimensional density  $q_k : D \rightarrow \mathbf{R}$  from the space  $L_2(D)$ , and  $k = 1, 2, \dots, m$ . The essence of (69) 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, for brevity the WTP algorithm, which is defined in Wasilkowski and Woźniakowski (1999), can be used to approximate linear tensor product operators, see the appendix. For the conditional expectation problem, the WTP algorithm is a quadrature formula of the form (47) which approximates  $E_k(V)/\|E_k\|$ .

Observe that we now have

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

Observe that  $\sup_d \sum_{j=1}^{\infty} \gamma_{d,kj} < \infty$  implies that all  $\|E_k\|$  are of order 1. Hence, the WTP algorithm can be also used to approximate  $E_k(V)$  as effectively as  $E_k(V)/\|E_k\|$ .

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  $\delta$  there exists a positive  $C_\delta$  and there is a WTP algorithm which is a quadrature formula of the form (47) for which the error bound (48) is

$$\varepsilon_j(V) = C_\delta j^{-1/p^*} \min_{c \in \mathbf{R}} \|V - c\|_{F_d}, \quad \text{with} \quad p^* = \max \left( p + \delta, \frac{2p_\gamma}{1 - p_\gamma} \right). \quad (70)$$

Here,  $p$  is the exponent of  $\varepsilon^{-1}$  for the one dimensional case,  $d = 1$ . We stress that neither  $C_\delta$  nor  $p^*$  depend on  $d$ . In particular, if

$$p_\gamma \leq \frac{p}{p+2} \quad (71)$$

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  $\varepsilon^{-1}$ . This with Theorems 3 and 4 yield

**Theorem 5** . Consider the spaces  $F_d$  with the reproducing kernel (67) and the weights  $\gamma_{d,k}$  satisfying (71). Then the conditional expectation problem is strongly tractable with strong exponent  $p$  which is the exponent of  $\varepsilon^{-1}$  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 quasi linear contraction problem is strongly tractable. The algorithm  $A$  defined by (49) with the WTP algorithm as  $\hat{E}_{k,j}$  computes an  $\varepsilon$ -approximation and its cost satisfies

$$\text{cost}(A) = O\left(c(d) m \alpha \varepsilon^{-p^*} \ln^{p^*} \varepsilon^{-1} + c(f) \alpha \varepsilon^{-p^*} \ln^{p^*+1} \varepsilon^{-1} + m \alpha^2 \varepsilon^{-2p^*} \ln^{2p^*} \varepsilon^{-1}\right), \quad (72)$$

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  $\beta$ .

We now present the results on strong tractability for the weighted Sobolev space  $F_{d,\gamma}$  defined in the appendix. It is known that the exponent  $p^*$  which appears in Theorem 5 is 1, see Novak (1998).

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

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

The maximum of this function is attained for  $x = [R, R, \dots, R]$  and  $K$  given by (66) is  $K = \prod_{k=1}^d (1 + R\gamma_{d,k})$ . 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  $p_\gamma \leq 1$ . Then the conditional expectation problem for any Markov densities is strongly tractable and the strong exponent is at most 2. With the additional assumption (71) that  $p_\gamma \leq 1/3$  we can apply Theorem 5 and the strong exponent is 1.

We add that it is known that the assumption  $p_\gamma \leq 1/3$  is *not* sharp for the Markov transition density  $p_k \equiv 1$ . It is proved in Hickernell and Woźniakowski (1999) that we can achieve the strong exponent of 1 assuming that  $p_\gamma \leq 1/2$ . The proof is, however, not constructive.

We now turn to strong tractability of the quasi linear contraction problem. This holds under the additional 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 (20) and (21) we now have

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

and the solution is

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

Observe that the integrals  $E_k(V)$  in (14) as well as quadrature formulas  $\hat{E}_{k,j}$  in (47) do not now depend on  $k$  and  $s$ . Hence,  $\hat{E}_{k,j}(V) = \hat{E}_j(V)$ , and the iteration (49) 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  $\pi$  is defined by (19). 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

$$\min_{c \in \mathbf{R}} \|V_i - c\|_{F_d} \equiv \min_{c \in \mathbf{R}} \|V_1 - c\|_{F_d} \leq \|V^*\|_{F_d}.$$

This proves that Theorem 5 holds with  $M = 1$ .

Note 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 Theorem 5 holds without the logarithms of  $\ln \varepsilon^{-1}$ .

## 6 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 quasi linear 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 finitely many function values of  $V$ . That is,  $V(s)$  is approximated by the linear algorithm<sup>10</sup>

$$\hat{V}_j(s) = \sum_{i=1}^j b_{i,j}(s)V(t_{i,j}) \tag{73}$$

---

<sup>10</sup>It is known that more general algorithms such as nonlinear algorithms using adaptive choice of sample points are not better than non-adaptive choice of sample points and linear algorithms considered in this section, see Traub, Wasilkowski and Woźniakowski (1988) and Traub and Traub and Werschulz (1998).

for some sample points  $t_{i,j} \in S_d$  and some functions  $b_{i,j}$  From the space  $L_2(S_d)$  with the norm  $\|f\|_{L_2(S_d)}^2 = \int_{S_d} f^2(t) dt$ . Let

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

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

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

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

$$n_{\text{app}}(\varepsilon, d) \leq C \varepsilon^{-p}, \quad \forall \varepsilon \in (0, 1), d = 1, 2, \dots \quad (74)$$

The smallest (or infimum of) such  $p$  is called the *strong exponent* of approximation.<sup>11</sup>

We now relate strong tractability of approximation to strong tractability of the conditional expectation and quasi linear contraction problems. Assume that (74) holds with  $C = C_{\text{app}}$  and  $p = p_{\text{app}}$ . 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

$$e(\hat{V}_j) \leq C_{\text{app}}^{1/p_{\text{app}}} j^{-1/p_{\text{app}}} \|V\|_{F_d}, \quad \forall V \in F_d.$$

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

$$\hat{E}_{k,j}(V)(s) = \int_{S_d} \hat{V}_j(t) p_k(t|s) dt = \sum_{i=1}^j \left( \int_{S_d} b_{i,j}(t) p_k(t|s) dt \right) V(t_{i,j}). \quad (75)$$

Hence,  $\hat{E}_{k,j}$  is of the form (47) with

$$a_{i,j,k}(s) = \int_{S_d} b_{i,j}(t) p_k(t|s) dt \quad \text{and} \quad t_{i,j,k} = t_{i,j}. \quad (76)$$

---

<sup>11</sup>There are a number of papers where strong tractability of approximation in various classes of function is considered, see e.g., Woźniakowski (1994), and Wasilkowski and Woźniakowski (1999), and a survey can be found in Traub and Werschulz (1998). 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 such as arbitrary linear functionals but we do not pursue this point here.

Note that 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 (53) with  $t_{p,j_{i-1},k} = t_{p,j_{i-1}}$ . Clearly,

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

and therefore

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

Let

$$P = \sup_{k \in [1,m], s \in S_d} \|p_k(\cdot|s)\|_{L_2(S_d)}. \quad (77)$$

If  $P$  is finite then the upper bound  $\varepsilon_j(V)$  of the quadrature error given by (48) is given by

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

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

$$j = C_{\text{app}} P^{p_{\text{app}}} \varepsilon^{-p_{\text{app}}}.$$

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

**Theorem 6** . *If  $P$  given by (77) 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 (60) is finite then the quasi linear 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 the strong exponent of approximation.*

*Furthermore, the algorithm  $A$  defined by (49) with  $\hat{E}_{k,j}$  given by (75) computes an  $\varepsilon$ -approximation at cost*

$$\text{cost}(A) = O\left(c(d) m \alpha \varepsilon^{-p_{\text{app}}} \ln^{p_{\text{app}}} \varepsilon^{-1} + c(f) \alpha \varepsilon^{-p_{\text{app}}} \ln^{p_{\text{app}}+1} \varepsilon^{-1} + m \alpha^2 \varepsilon^{-2p_{\text{app}}} \ln^{2p_{\text{app}}} \varepsilon^{-1}\right)$$

*with  $\alpha = C_{\text{app}}(MP)^{p_{\text{app}}}$  and the big  $O$ -factor depending only on  $\beta$ .*

We illustrate Theorem 6 for the space  $F_d$  which was already considered in Section 5. This is the Hilbert weighted tensor product space of functions defined on  $S_d = D^d$  with the reproducing kernel  $K_d$  given by (67). We assume that

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

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

$$\lambda_1 \geq \lambda_2 \geq \dots \geq 0.$$

Let  $p_\lambda$  be the sum-exponent of the sequence  $\{\lambda_i\}$  of eigenvalues defined by (68). As in Section 5,  $p_\gamma$  is the sum-exponent of the sequence  $\{\gamma_{d,k}\}$  of weights. It is proven in Wasilkowski and Woźniakowski (1999) that the necessary condition on strong tractability of approximation in  $F_d$  is that both  $p_\lambda$  and  $p_\gamma$  are finite, and then the strong exponent of approximation is at least  $2 \max(p_\lambda, p_\gamma)$ .

Assume then that  $p_\lambda$  and  $p_\gamma$  are finite. In this case, the WTP algorithm is also effective for approximation, see Wasilkowski and Woźniakowski (1999). 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 (73) 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  $k$ th component of the vector  $s$ . If we assume that

$$p_\gamma \leq \frac{p_1}{2 + 2p_1} \tag{78}$$

then the WTP algorithm computes an  $\varepsilon$ -approximation at cost  $C_\delta \varepsilon^{p_1 + \delta}$ . Here,  $\delta$  is positive and can be made arbitrarily small and  $C_\delta$  is independent of  $d$  and may only depend on  $\delta$ . 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 6 yield

**Corollary 2** . *Consider the spaces  $F_d$  with the reproducing kernel (67) and the weights  $\gamma_{d,k}$  satisfying (78). Then approximation in  $F_d$  is strongly tractable with strong exponent at most  $p_1$ , where  $p_1$  is the exponent of  $\varepsilon^{-1}$  for the univariate case.*

*If  $P$  given by (77) 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 (60) is finite then the quasi linear 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 (49) with the WTP algorithm to obtain  $\hat{E}_{k,j}$  by (75) computes an  $\varepsilon$ -approximation and its cost satisfies

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

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

As in the previous section, we now specify the results for the weighted Sobolev space  $F_{d,\gamma}$  defined in the appendix. We now have  $p_1 = 1$  and (78) means that  $p_\gamma \leq 1/4$ .

Consider the case when the transition densities are the same  $p_k(t|s) = p(t|s)$ . Due to (19), the quasi linear contraction problem takes now the form

$$V(s) = \pi(s) + \beta \int_{[0,R]^d} V(t)p(t|s) dt, \quad \pi(s) = f(\pi_1(s), \dots, \pi_m(s)). \quad (79)$$

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

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

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

$$e(\hat{V}_j) \leq \min\left(Cj^{-1+\delta}, c\right) \|V\|_{F_{d,\gamma}}, \quad (80)$$

see Theorem 5 of Wasilkowski and Woźniakowski (1999) applied to the problem of approximating functions  $V$  from  $F_{d,\gamma}$ . The need of the constant  $c$  will be soon clear.

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

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

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

We now estimate the ratios  $\|V_i\|_{F_{d,\gamma}}/\max(\|V^*\|, \|V^*\|_{F_{d,\gamma}})$ . For any  $u \in \{1, 2, \dots, d\}$  by  $V^u$  we mean  $V(0)$  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

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

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

$$|E^u(V)(s) - \hat{E}_j^u(V)(s)| \leq e(\hat{V}_j) \|p^u(\cdot|s)\|_{L_2([0,R]^d)}$$

and

$$\|E(V) - \hat{E}_j(V)\|_{F_{d,\gamma}} \leq e(\hat{V}_j) \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}},$$

where

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

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

$$\|V_i - V^*\|_{F_{d,\gamma}} \leq \beta \left( e((\hat{V}_{i-1})_{j_{i-1}}) + \|V_{i-1} - V^*\| \right) \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}. \quad (81)$$

From (80) we have  $e(\hat{V}_j) \leq c(\|V - V^*\|_{F_{d,\gamma}} + \|V^*\|_{F_{d,\gamma}})$ . Since  $\|V_{i-1} - V^*\|$  is of order  $\|V^*\|$ , we have

$$\|V_i - V^*\|_{F_{d,\gamma}} \leq \beta c \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}} \|V_{i-1} - V^*\|_{F_{d,\gamma}} + K \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}} \|V^*\|$$

for some  $K$  dependent only on  $\beta$ .

Choose  $c$  such that  $\beta c \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}} \leq 1/2$ , say. Observe that as long as  $\|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}$  is uniformly bounded in  $d$  then  $c$  is uniformly bounded from below and the presence of  $c$  in (80) is not really essential. From this we have

$$\|V_i - V^*\|_{F_{d,\gamma}} = O\left(\max(\|V^*\|_{F_{d,\gamma}}, \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}} \|V^*\|)\right).$$

This proves that  $\|V_i\|_{F_{d,\gamma}}/\max(\|V^*\|, \|V^*\|_{F_{d,\gamma}})$  is of order 1 as long as  $\|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}$  is uniformly bounded in  $d$ .

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

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

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

$$U = \{ (\pi_1, \pi_2, \dots, \pi_m) : f(\pi_1(\cdot), \pi_2(\cdot), \dots, \pi_m(\cdot)) \in F_{d, \gamma} \}.$$

Then Theorem 5 and the results of this section yield

**Theorem 7 .** *If  $p_\gamma \leq 1/4$  then the conditional expectation and quasi linear contraction problems with data from  $P$  and  $U$  are strongly tractable with strong exponent 1. For any positive  $\delta$ , the algorithm  $A$  defined by (49) with an appropriately chosen WTP algorithm computes an  $\varepsilon$ -approximation with*

$$\begin{aligned} \text{cost}(A) = O \left( c(d) m L^{2(1+\delta)} \varepsilon^{-1-\delta} \ln^{1+\delta} \varepsilon^{-1} \right. &+ c(f) m L^{2(1+\delta)} \varepsilon^{-1-\delta} \ln^{2+\delta} \varepsilon^{-1} \\ &+ \left. m L^{4(1+\delta)} \varepsilon^{-2(1+\delta)} \ln^{2(1+\delta)} \varepsilon^{-1} \right) \end{aligned}$$

with the big  $O$ -factor depending only on  $\beta$ ,  $\delta$  and  $R$ .

## 7 Example: The Rational Expectations Pricing Model

In this section we provide an example of how special structure can arise in an economic problem. We show how this structure enables us to design an algorithm that nearly attains the optimal univariate rate of convergence of  $n^{-1}$ , in multivariate problems where the dimension  $d$  may be arbitrarily large. We consider a special case of the rational expectations asset pricing model, where  $V^*(s)$  denotes the expected discounted value of an asset given information  $s$  and is the unique solution to the Fredholm integral equation given by

$$V(s) = \pi(s) + \beta \int_{[0, R]^d} V(t) p(t|s) dt, \quad s \in [0, R]^d, \quad (82)$$

where

$$p(t|s) = \prod_{i=1}^d \frac{\exp(-(t_i - \alpha_{d,i} - b_d s_i)^2 / (2\lambda_{d,i}))}{\int_0^R \exp(-(t - \alpha_{d,i} - b_d s_i)^2 / (2\lambda_{d,i})) dt}, \quad (83)$$



with  $\alpha_d = Q_d a_d + (1 - b_d)/2\vec{R}$ , where  $Q_d$  is a  $d \times d$  orthogonal matrix,  $\lambda_{d,i} > 0$ ,  $a_d$  is an element of  $R^d$ ,  $b_d \in (0, 1)$  for all  $d$ , and  $\vec{R} = [R, \dots, R]$ . The interpretation of this problem is that  $\pi(s)$  represents the payoff of the asset in state  $s$ , and  $p(t|s)$  is the Markov transition probability governing the evolution of information. The transition density (83) is a truncated normal approximation to the vector  $AR(1)$  process given by:

$$\ln(s_{j+1}) = a_d + b_d \ln(s_j) + \varepsilon_j \quad (84)$$

where  $\{\varepsilon_j\}$  is an *IID* Gaussian process with marginal distribution  $N(0, \Omega_d)$  where  $\Omega_d$  is a positive definite covariance matrix. We can write  $\Omega_d = Q_d' D_d Q_d$ , where  $Q_d$  is an orthogonal matrix and  $D_d = \text{diag}(\lambda_{d,1}, \dots, \lambda_{d,d})$  is a diagonal matrix containing the eigenvalues of  $\Omega_d$ . Formula (83) results from truncating the normal process to the cube  $[0, R]^d$ .

We consider the Fredholm integral problem (82) for the weighted Sobolev space  $F_d = F_{d,\gamma}$  which is defined in the appendix. We take the vector  $\gamma$  of the weights given by  $\gamma_{d,i} = i^{-4}$ .

**Theorem 8** . Assume that  $\pi$  belongs to  $F_{d,\gamma}$ ,

$$b := \sup_d b_d < 1 \quad \text{and} \quad a := \sup_d \left( \sum_{i=1}^d a_{d,i}^2 \right)^{1/2} < \infty$$

and

$$\lambda := \sup_d \sum_{i=1}^d i^4 \lambda_{d,i}^{-2} < \infty.$$

Then there exists  $R_0$  depending only on  $a$  and  $b$  such that the rational expectations pricing model (82) with  $R \geq R_0$  is strongly tractable. More precisely, for any positive  $\delta$  there exists a positive  $C$  depending only on  $\delta, a, b, \lambda$  and  $R$  such that the algorithm  $A$  computes an  $\varepsilon$ -approximation with

$$\text{cost}(A) \leq C \left( c(d) \varepsilon^{-(1+\delta)} \ln^{1+\delta} \varepsilon^{-1} + \varepsilon^{-(2+\delta)} \ln^{2(1+\delta)} \varepsilon^{-1} \right).$$

Hence, the strong exponent of the rational expectations pricing model is at most 1, and the strong exponent of arithmetic operations is at most 2.

The proof of Theorem 8 is presented in the appendix. Notice that if condition (8) holds the parameters  $\{\lambda_{d,i}\}$  tend to  $\infty$  as  $i \rightarrow d$  and  $d \rightarrow \infty$ . Formula (83) implies that for  $i$  close to  $d$ , the corresponding marginal densities of  $p(t|s)$  are close to a uniform density over  $[0, R]$  independent of the value  $s_i$ . This implies that even though  $\pi(s)$  depends on potentially all

of the state variables  $s$ , if it belongs to the weighted Sobolev space  $F_{d,\gamma}$  then the conditional expectation  $E\pi$  and the asset value  $V^*$  will only effectively depend on only a relatively small number of the  $s_i$  variables. The economic interpretation of the special structure embodied by this example is that in problems where there sufficient uncertainty about the future values of many of the state variables affecting asset payoffs, asset prices will effectively depend only on a finite number of state variables for which there is a sufficiently strong link between current realized values and future expectations. The implied restrictions on the dependence of  $V^*$  on  $s$  makes the calculation of asset prices strongly tractable, and the number of sample points  $n$  necessary to compute an  $\varepsilon$ -approximation to  $V^*(s)$  is independent of  $d$ .

Observe that we did not assume that the norm in the space  $F_{d,\gamma}$  of the function  $\pi$  is uniformly bounded in  $d$ . However, the norm of  $\pi$  directly affects the norm of the solution  $V^*$  which, in turn, affects the definition of an  $\varepsilon$ -approximation.

## 8 Conclusion

In this paper we have identified a general type of “special structure” and have introduced an algorithm that enables us to exploit this special structure and break the curse of dimensionality associated with approximating the fixed point  $V = \Gamma(V)$  for a class of quasilinear contraction mappings  $\Gamma$  that arise frequently in economic applications. We showed that there is a deterministic, successive approximations algorithm which converges at a faster rate than the random successive approximations and random multigrid algorithms that were used by Rust (1997) to break the curse of dimensionality for these problems in the randomized setting. In addition to showing that a deterministic algorithm can break the curse of dimensionality, we have established the surprising result that even though the function  $V$  can depend on an arbitrarily large number of continuous-valued arguments  $(s_1, \dots, s_d)$ , our algorithm can approximate  $V$  to within an error of  $\varepsilon$  using only roughly  $O(\varepsilon^{-1})$  function evaluations and  $O(\varepsilon^{-2})$  arithmetic operations, *independent of  $d$* . In the terminology of computer science, we have shown that the quasilinear contraction problem is strongly tractable with strong exponent equal to 1: using  $n$  function evaluations and  $O(n^2)$  arithmetic operations, our algorithm produces an approximation to the true fixed point  $V$  with an error bounded by  $Cn^{-p}$ , where  $p$  is close to one and  $C$  is an absolute constant that does not depend on  $d$ . Thus, we have identified a class of multivariate problems with arbitrarily large  $d$  for which our algorithm attains nearly the same rate of convergence that can be attained for the uni-dimensional problems,  $d = 1$ . However, for multi-dimensional problems with  $d$  continuous variables, the method of successive approximations using multi-dimensional

interpolation and standard product-rule quadratures to perform the numerical integrations requires  $O(n^d)$  points and  $O(n^{2d})$  arithmetic operations to attain an error of order  $O(1/n)$ . For these methods the curse of dimensionality is present. We use an alternative linear approximation algorithm known as the *weighted tensor product algorithm* to achieve an error roughly of  $C/n$  using only  $n$  points (function evaluations).

The other contribution of this paper is to identify an economically meaningful type of “special structure” for the profit, utility, or value functions entering the fixed point problem for which it is possible to nearly achieve this optimal uni-dimensional rate of convergence. The special structure can be described intuitively: it occurs when the dependence of a function on the  $i$ th variables decreases with increasing  $i$ . This dependence is controlled by a sequence of positive weights  $\gamma_i$ . We have provided a criterion on  $\gamma_i$  that enables us to break the curse of dimensionality. A sufficient condition for strong tractability is that the weights satisfy  $\sum_{i=1}^{\infty} \gamma_i < \infty$ . The boundedness of this sum implies that the weights  $\gamma_i$  go to zero. When  $\gamma_i$  is small, the norm of the function is very sensitive to changes in the  $i$ th variable. Thus, if our functions are to have bounded weighted norm, they must essentially be “flat” with respect to variables with large  $i$ . If we require that the  $\gamma_i$  weights approach zero at a sufficiently rapid rate, so that the sum of the weights satisfies  $\sum_{i=1}^{\infty} \gamma_i^{1/4} < \infty$ , then the strong exponent of our algorithm is  $p = 1$ , i.e., it comes arbitrarily close to attaining the optimal uni-dimensional rate of convergence rate of  $Cn^{-p}$  with  $p = 1$ .

We note two important caveats about our results. First, although our algorithm is constructive, we do not yet feel that it is “practical” for use in real problems. The weighted tensor product algorithm is difficult to implement which may imply that its advantages over simpler algorithms such as Rust’s (1997) random multigrid method will only become obvious for large  $d$ . A related problem is that our analysis also assumes that the quadrature weights that are used to approximate the conditional expectations of the value functions are “precomputed” (see formula (76) in Section 6). However, computing these weights themselves involve computing multivariate integrals. Our analysis has assumed that these integrals are computed exactly, but in practice they would have to be computed numerically and this would be a substantial additional computational burden.

Second, although the Bellman equation of dynamic programming is a special type of quasi-linear contraction mapping, we do not yet know whether our result applies to this case. The reason is that our result requires sufficient smoothness of both the profit/utility functions  $\{\pi_k\}$  and the sequence of value functions  $\{V_j\}$  generated by our successive approximations algorithm. The assumption in Theorem 7 that  $f(\pi_1(\cdot), \dots, \pi_m(\cdot)) \in F_{d,\gamma}$  does not generally hold when the quasi linear function  $f$  is non-differentiable as in the Bellman case where  $f$  is the max function. Further, the  $\{V_j\}$  sequence must have bounded weighted norm for strong tractability to hold and the strong exponent to be equal to 1. However, in the Bellman case,

the max operator may introduce enough kinks in the  $\{V_j\}$  functions so that the weighted norm of this sequence may not be uniformly bounded. One way to get around the problem is by “adding some noise” to smooth out the kinks: thus, the “smoothed Bellman operators”  $\Gamma_\sigma$  with smoothing parameters  $\sigma > 0$  do have sufficient smoothness for our results to apply provided  $\sigma$  is sufficiently large. However, as  $\sigma \rightarrow 0$ , the norm of the corresponding smoothed quasi-linear function  $f_\sigma(\pi_1(\cdot), \dots, \pi_m(\cdot))$  tends to infinity, and for this reason we cannot use the results for the strong tractability of the smoothed Bellman problem to argue that the original unsmoothed Bellman problem is also strongly tractable.

We are optimistic that it will be possible to extend our results to the important case of Bellman operators, but we leave this as an open conjecture and a topic for future research. One possible direction is to consider policy iteration algorithms for solving the Bellman equation and to note that each policy valuation step involves a solution to a contractive Fredholm integral equation, and our results have shown that under appropriate conditions, the contractive Fredholm problem is strongly tractable with strong exponent equal to 1. We conclude by noting that our theoretical results have already stimulated new investigations that attempt to use linear algorithms similar to the WTP algorithm to break the curse of dimensionality of dynamic programming problems. Computational experiments such as in Benítez-Silva *et. al.* (2000) suggests that the strategy of using linear algorithms in conjunction with policy iteration could be highly effective for solving high dimensional dynamic programming problems.

## 9 Appendix

### 9.1 Weighted Sobolev space $F_{d,\gamma}$ .

The space  $F_{d,\gamma}$  is a Hilbert space of functions defined on the  $d$ -dimensional cube  $S_d = [0, R]^d$ , see Sloan and Woźniakowski (1998) where the case  $R = 1$  is considered. This is a Hilbert space with a *reproducing kernel*. For the basic theory of such spaces the reader is referred to e.g., Aronszajn (1950) and Wahba (1990). The most important property of a reproducing kernel Hilbert space is that there exists a function  $K_d : S_d \times S_d \rightarrow R$  such that  $K_d(\cdot, x) \in H$  for any  $x \in S_d$  and

$$f(x) = \langle f, K_d(\cdot, x) \rangle \quad \forall f \in H,$$

where  $\langle \cdot, \cdot \rangle$  is the inner product of  $H$ .

The function  $K_d$  is called a *reproducing kernel* of the Hilbert space  $H$ . Sometimes we write  $H = H(K_d)$  to indicate the reproducing kernel of  $H$ . The reproducing kernel has the

following properties. For any  $x$  and  $t$  from  $S_d$ , we have

$$\|K_d(\cdot, x)\| = K_d^{1/2}(x, x) \quad \text{and} \quad K_d(t, x) \leq K_d^{1/2}(t, t)K_d^{1/2}(x, x).$$

For any integer  $n$  and points  $t_i \in S_d$  for  $i = 1, \dots, n$ , the  $n \times n$  matrix  $(K_d(t_i, t_j))$  is non-negative definite. In fact, for any function  $K_d$  with the last property there exists a Hilbert space for which  $K_d$  is its reproducing kernel. Therefore, it is enough to present a reproducing kernel in order to define the Hilbert space with this kernel.

That is the approach we take for introducing the weighted Sobolev space  $F_{d,\gamma}$ . Its reproducing kernel  $K_{d,\gamma}$  is

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

Here,  $\gamma_{d,i} \geq 0$ . For  $\gamma_{d,i} \equiv 1$ , we obtain the reproducing kernel of the classical (unweighted) Sobolev space  $F_{d,1} = W^{1,1,\dots,1}([0, R]^d)$ .

For  $d = 1$ , the space  $F_{1,\gamma}$  consists of absolutely continuous functions whose first derivatives are in  $L_2([0, 1])$  with the inner product of  $V, W \in F_{1,\gamma}$  given by<sup>12</sup>

$$\langle V, W \rangle = V(0)W(0) + \gamma_{1,1}^{-1} \int_0^1 V'(x)W'(x) dx.$$

For  $d \geq 2$ , the space  $F_{d,\gamma}$  is the tensor product of  $F_{1,\gamma_{d,1}} \otimes F_{1,\gamma_{d,2}} \otimes \dots \otimes F_{1,\gamma_{d,d}}$  and corresponds to functions which are once differentiable with respect to each variable.

We denote the inner product and norm in  $F_{d,\gamma}$  by  $\langle \cdot, \cdot \rangle_{F_{d,\gamma}}$  and  $\|\cdot\|_{F_{d,\gamma}} = \langle \cdot, \cdot \rangle_{F_{d,\gamma}}^{1/2}$ . The inner product of  $F_{d,\gamma}$  is

$$\langle V, W \rangle = \sum_{u \subset \{1,2,\dots,d\}} \gamma_{d,u}^{-1} \int_{[0,R]^{|u|}} \frac{\partial^{|u|}}{\partial x_u} V(x_u, 0) \frac{\partial^{|u|}}{\partial x_u} W(x_u, 0) dx_u. \quad (85)$$

---

<sup>12</sup>For the inner product

$$\langle V, W \rangle = \int_0^1 V(x)W(x) dx + \gamma^{-1} \int_0^1 V'(x)W'(x) dx.$$

we obtain the Sobolev space with the reproducing kernel

$$K_1(x, t) = \frac{\sqrt{\gamma}}{\sinh \sqrt{\gamma}} \cosh(\sqrt{\gamma}(1 - \max(x, t))) \cosh(\sqrt{\gamma} \min(x, t)), \quad \forall x, t \in [0, 1].$$

as shown by Thomas-Agnan (1996).

Here,  $|u|$  is the cardinality of  $u$ . For the vector  $x \in [0, R]^d$ , we denote  $x_u$  as the vector From  $[0, R]^{|u|}$  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  $[0, R]^d$ , 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  $j$ th variable, and we have  $0/0 = 0$  in the inner product formula. Observe that the sum in the inner product has  $2^d$  terms.

The norms  $\|V\|_{F_{d,\gamma}}$  and  $\|V\| = \max_{x \in [0,R]^d} |V(x)|$  may be quite different. Indeed, take  $R = 1$  and the function  $V(x) = (1 - x_1) \dots (1 - x_d)$ . Then  $\|V\| = 1$  and for positive  $\gamma_{d,j}$  we have

$$\|V\|_{F_{d,\gamma}} = \sum_u \gamma_u^{-1} = \prod_{j=1}^d (1 + \gamma_{d,j}^{-1}).$$

Hence, for  $\gamma_{d,j} \equiv 1$  we have  $\|V\|_{F_{d,\gamma}} = 2^d$ .

We now consider the Cobb-Douglas function,

$$\pi(x) = \prod_{k=1}^d (x_k + a_k)^{\alpha_k}, \quad x_k \in [0, R],$$

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

The weighted norm of  $\pi$  was estimated in Wasilkowski and Woźniakowski (1999) for  $\gamma_{d,k} = \alpha_k$ . Then  $\pi \in F_{d,\gamma}$  iff  $a := \min_j a_j > 0$ . If so then

$$\|\pi\|_{F_{d,\gamma}} \leq b^2 + \frac{b^2 R}{a^2} \exp(R/a^2), \quad \forall d, \quad (86)$$

where  $b = \max_j a_j$ . In particular, for  $a_j \equiv 1$  we have  $\|\pi\|_{F_{d,\gamma}} \leq 1 + R \exp(R)$ .

## 9.2 Weighted Tensor Product Algorithm

The WTP algorithm is defined as a linear algorithm to approximate linear multivariate weighted tensor product problems, see Wasilkowski and Woźniakowski (1999). The essence of this algorithm is that it requires only the knowledge of linear algorithms for the solution of the corresponding univariate problem. This is usually relatively easy to achieve. For the multivariate case, the WTP algorithm takes a special tensor product of the known univariate algorithms in such a way that the total number of information and arithmetic operations

is strictly controlled. This number depends on the weights of the tensor product problem. For the weights which go sufficiently fast to zero, the total number of information and arithmetic operations needed to guarantee that the error of the WTP algorithm is at most  $\varepsilon$  is independent of the dimension  $d$  and, roughly speaking, is the same as for  $d = 1$ .

We briefly describe the WTP algorithm for multivariate integration

$$I(f) = \int_{S_d} f(x) dx.$$

Let  $\{I_{n,\gamma}\}$  be a sequence of algorithms of the form (27) for approximation of the integral  $I(f)$  in the univariate case  $d = 1$ . That is,

$$I_{n,\gamma}(f) = \sum_{i=1}^n w_{n,i,\gamma} f(s_{n,i,\gamma}) \quad (87)$$

where the quadrature weights  $\{w_{n,i,\gamma}\}$  and sample points  $\{s_{n,i,\gamma}\}$  may depend on the weight  $\gamma$  defining the weighted norm for  $F_{1,\gamma}$ . Observe that the cost of  $I_{n,\gamma}$  is  $c(d)n$ . We assume that the quadrature weights and sample points are chosen so that the errors,  $e_n(F_{d,\gamma})$  converge to 0 as  $n \rightarrow \infty$ . For each weight  $\gamma$ , we assume that there is an increasing sequence of integers

$$m_{0,\gamma} = 0 < m_{1,\gamma} = 1 < m_{2,\gamma} < \dots < m_{i,\gamma}, \quad (88)$$

and define

$$\Delta_{i,\gamma}(f) = I_{m_{i,\gamma},\gamma}(f) - I_{m_{i-1,\gamma},\gamma}(f) \quad \text{for } i \geq 1. \quad (89)$$

Observe that  $\sum_{i=1}^j \Delta_{i,\gamma} = I_{m_{j,\gamma},\gamma}$  and  $I_{m_{j,\gamma},\gamma}(f)$  converges to  $I(f)$  for every  $f \in F_{d,\gamma}$ . Let  $\mathbb{N}_+^d$  be the set of vectors  $\vec{i} = [i_1, \dots, i_d]$  with positive integer coefficients  $i_k$ . To stress their role, we shall refer to them as multi-indices. By  $|\vec{i}|$  we mean  $\sum_{k=1}^d i_k$ . Let  $\{P_{n,d}\}$  be a sequence of subsets of  $\mathbb{N}_+^d$  such that  $P_{n,d}$  consists of  $n$  multi-indices,  $P_{n,d} \subset P_{n+1,d}$  and  $\bigcup_n P_{n,d} = \mathbb{N}_+^d$ . Each set  $P_{n,d}$  may depend on all weights  $\gamma_i$  for  $i = 1, 2, \dots, d$ .

The *weighted tensor product (WTP)* algorithm is defined as the sequence  $\{U_{n,d,\gamma}\}$  given by

$$U_{n,d,\gamma}(f) = \sum_{\vec{i} \in P_{n,d}} \left( \bigotimes_{k=1}^d \Delta_{i_k, \gamma_k} \right) (f), \quad (90)$$

where the *tensor product*  $f = f_1 \otimes \dots \otimes f_d = \bigotimes_{k=1}^d f_k$  in the case of scalars  $f_k$  is just the product  $\prod_{k=1}^d f_k$ . In the case where the  $f_k$  are scalar functions,  $f = \bigotimes_{k=1}^d f_k$  is a function of  $d$  variables given by  $f(t_1, \dots, t_d) = \prod_{k=1}^d f_k(t_k)$ . In the case of linear operators  $T_k$ ,  $T = \bigotimes_{k=1}^d T_k$

is a linear operator such that  $T\left(\bigotimes_{k=1}^d f_k\right) = \bigotimes_{k=1}^d T_k(f_k)$ . The WTP algorithm depends on a number of parameters. First of all, it depends on the sequence of weights  $\{\gamma_{d,k}\}$ , the sequence of cardinalities  $m_{i,\gamma_k}$ , as well as the sequence of sets  $P_{n,d}$ . It also depends on the one-dimensional quadrature algorithms  $I_{n,\gamma_{d,k}}$ . By varying these parameters we obtain the class of WTP algorithms.

Since  $\lim_{j \rightarrow \infty} \sum_{i=1}^j \Delta_{m_{i,\gamma}}(f) = I(f)$  for every  $f \in F_{d,\gamma}$ , we have

$$I(f) = \sum_{\vec{i} \in \mathbf{N}_+^d} \left( \bigotimes_{k=1}^d \Delta_{i_k, \gamma_{d,k}} \right) (f), \quad \forall f \in F_{d,\gamma} \quad (91)$$

This yields

$$|I(f) - U_{n,d,\gamma}(f)| = \left\| \sum_{\vec{i} \in \mathbf{N}_+^d \setminus P_{n,d}} \bigotimes_{k=1}^d \Delta_{i_k, \gamma_{d,k}}(f) \right\| \leq \sum_{\vec{i} \in \mathbf{N}_+^d \setminus P_{n,d}} \left\| \bigotimes_{k=1}^d \Delta_{i_k, \gamma_{d,k}}(f) \right\|.$$

Therefore the error of  $U_{n,d,\gamma}$  is bounded by

$$e(U_{n,d,\gamma}, I) \leq \sum_{\vec{i} \in \mathbf{N}_+^d \setminus P_{n,d}} \left\| \bigotimes_{k=1}^d \Delta_{i_k, \gamma_{d,k}} \right\| = \sum_{\vec{i} \in \mathbf{N}_+^d \setminus P_{n,d}} \prod_{k=1}^d \left\| \Delta_{i_k, \gamma_{d,k}} \right\|. \quad (92)$$

This formula suggests that a good choice for  $P_{n,d}$  is the set of  $n$  multi-indices  $\vec{i}$  which correspond to the  $n$  largest norms of  $\bigotimes_{k=1}^d \Delta_{i_k, \gamma_{d,k}}$ . We refer the reader to Wasilkowski and Woźniakowski (1999) for further discussion of the WTP and its rate of convergence.

### 9.3 Proof of Theorem 8

We prove Theorem 8 by applying Theorem 7. First of all, note that for  $\gamma_{d,i} = i^{-4}$  we have  $p_\gamma = 1/4$  as needed in Theorem 7. We now check the other two assumptions of Theorem 7 that

$$\|p(\cdot|s)\|_{L_2([0,R]^d)} \quad \text{and} \quad \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}$$

are uniformly bounded in  $d$  and  $s$ .

To prove that  $\|p(\cdot|s)\|_{L_2([0,R]^d)}$  is uniformly bounded in  $d$  and  $s$ , observe that due to (83) and the conditions on  $\lambda_{d,i}$ , it is enough to show that for  $\lambda \geq 1$  and  $c = \alpha_{d,i} + b_d s_i$  we have

$$\int_0^R \exp\left(-\frac{(u-c)^2}{\lambda}\right) du \leq \left( \int_0^R \exp\left(-\frac{(u-c)^2}{(2\lambda)}\right) du \right)^2.$$



This holds if we show that

$$\int_0^R \exp\left(-\frac{(u-c)^2}{2\lambda}\right) du \geq 1.$$

We have

$$-a + \frac{1-b}{2}R \leq c \leq a + \frac{R}{2} + b_d \left(s_i - \frac{R}{2}\right) \leq a + \frac{1+b}{2}R.$$

Therefore  $-c \leq a - (1-b)R/2$  and  $R - c \geq -a + (1-b)R/2$  and

$$\int_0^R \exp\left(-\frac{(u-c)^2}{2\lambda}\right) du = \int_{-c}^{R-c} \exp\left(-\frac{x^2}{2\lambda}\right) dx \geq \int_{a-(1-b)R/2}^{-a+(1-b)R/2} \exp\left(-\frac{x^2}{2}\right) dx.$$

Since the last integral is about  $\sqrt{2\pi} > 1$  for large  $R$ , there exists  $R_0$  depending only on  $a$  and  $b$  such that the last integral is indeed at least 1, as claimed.

We now estimate  $\|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}$ . Due to the product form of  $p$  in (83) we first consider the function

$$h(t_i|s_i) = \frac{\exp\left(-\frac{(t_i - \alpha_{d,i} - b_d s_i)^2}{2\lambda_{d,i}}\right)}{\int_0^R \exp\left(-\frac{(t - \alpha_{d,i} - b_d s_i)^2}{2\lambda_{d,i}}\right) dt}.$$

It is easy to check that

$$\int_0^R \int_0^R \left(\frac{\partial h}{\partial s}(t|s)\right)^2 dt ds = O\left(\frac{b_d^2}{\lambda_{d,i}^2}\right)$$

with the factor in the big  $O$ -notation depending on the global parameters  $a, b$  and  $R$ . This yields that

$$\begin{aligned} \|p\|_{L_2([0,R]^d) \times F_{d,\gamma}}^2 &= \|p(\cdot|0)\|_{L_2([0,R]^d)}^2 + \sum_{\emptyset \neq u \subset \{1, \dots, d\}} O\left(\prod_{j \in u} \frac{b_d^2}{\gamma_{d,i} \lambda_{d,i}}\right) \\ &= O\left(\prod_{j=1}^d \left(1 + \frac{i^4 b C}{\lambda_{d,i}}\right)\right) \end{aligned}$$

for some  $C$ . This is uniformly bounded in  $d$  due to the condition on  $\lambda_{d,i}$  which completes the proof.

## References

- [1] V. Aguirregabiria “The Dynamics of Markups and Inventories in Retailing Firms” *Review of Economic Studies* **66** 275–308, 1999.
- [2] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.*, **68**, 337–404, 1950.
- [3] N. S. Bakhvalov, On approximate calculation of integrals, *Vestnik MGU, Ser. Mat. Mekh. Astron. Fiz. Khim.*, **4**, 3–18. (In Russian.)
- [4] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, N.J., 1957.
- [5] H. Benítez-Silva, G. Hall, G. Hitsch, G. Pauletto, and J. Rust “Comparison of Discrete and Parametric Approximation Methods for Continuous State Dynamic Programming Problems” manuscript, Yale University.
- [6] D. Blackwell, Discounted Dynamic Programming, *Annals of Mathematical Statistics*, **36**, 226–235, 1965.
- [7] P. Boyle, M. Broadie, and P. Glasserman “Monte Carlo Methods for Security Pricing” *Journal of Economic Dynamics and Control* **8-9** 1267–1322, 1997.
- [8] C. Chow and J. Tsitsiklis, The complexity of dynamic programming, *J. of Complexity*, **5**, 466–488, 1989.
- [9] R. Cooper, *Dynamic Programming: Theory and Applications* manuscript, Boston University, 1999.
- [10] E. Denardo, Contraction Mappings Underlying the Theory of Dynamic Programming, *SIAM Review*, **9**, 165–177, 1967.
- [11] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics, **1651**, 1997.
- [12] D. Duffie *Security Markets: Stochastic Models*, Academic Press, 1988.
- [13] 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.

- [14] G. Hall and J. Rust “An Empirical Model of Inventory Investment by Durable Commodity Intermediaries” Carnegie-Rochester Conference Series on Public Policy **52** 171–214, 2000.
- [15] S. Heinrich, Monte Carlo Complexity of Global Solution to Integral Equations, *J. of Complexity*, **14-2**, 151–175, 1998.
- [16] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimensions, submitted for publication, 1998.
- [17] R. Howard, *Dynamic Programming and Markov Processes* New York, Wiley, 1960.
- [18] Z. Huang, L. Khachiyan and K. Sikorski, Approximating fixed points of weakly contractive mappings, to appear in *J. Complexity*, 1999.
- [19] K. Judd *Numerical Methods in Economics* MIT Press, Cambridge, 1998.
- [20] F. Kydland and E. Prescott “Time to Build and Aggregate Fluctuations” *Econometrica* **50** 1345–1370, 1982.
- [21] R.E. Lucas, Jr., Asset prices in an exchange economy, *Econometrica*, **46-6**, 1429–1446, 1978.
- [22] 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.
- [23] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM CBMS-NSF Monograph 63, Philadelphia, PA, 1992.
- [24] E. Novak, *Deterministic and stochastic error bounds in numerical analysis*, Lecture Notes in Mathematics, Springer Verlag, Berlin, 1349, 1988.
- [25] E. Novak and H. Woźniakowski, On the cost of uniform and nonuniform algorithms, *Th. Computer Science*, **219**, 301–318, 1999.
- [26] E. Novak and H. Woźniakowski, Intractability results for integration and discrepancy, submitted for publication, 1999.
- [27] A. Papageorgiou and J.F. Traub, Beating Monte Carlo *Risk*, 1996.

- [28] A. Papageorgiou and J.F. Traub, Faster evaluation of multivariate integrals, *Computations in Physics*, 1997.
- [29] S. H. Paskov, New methodologies for valuing derivatives, in “Mathematics of Derivative Securities”, ed. by S. Pliska and M. Dempster, Isaac Newton Institute, *Cambridge University Press*, Cambridge, UK, 1997.
- [30] S. Paskov and J. Traub, Faster valuation of financial derivatives, *Journal of Portfolio Management*, **22**, 113–120, 1995.
- [31] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes*, Second Edition, Cambridge University Press, 1992.
- [32] J. Rust, Stationary Equilibrium in a Market for Durable Assets, *Econometrica*, **53–4**, 783–805, 1985.
- [33] J. Rust, Structural Estimation of Markov Decision Processes, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, **4**, 3082–3139, North Holland, 1994.
- [34] J. Rust, Using randomization to break the curse of dimensionality, *Econometrica*, **65–3**, 487–516, 1997.
- [35] 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>).
- [36] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *J. Complexity*, **14–1**, 1–33, 1998.
- [37] I. H. Sloan and H. Woźniakowski, Tractability of deterministic and Monte Carlo integration algorithms in non-periodic and periodic weighted tensor product Hilbert spaces, submitted for publication, 2000.
- [38] K. Sikorski and H. Woźniakowski, Complexity of fixed points I, *J. Complexity*, 388–405, **3**, 1987.
- [39] N. Stokey and R. Lucas *Recursive Methods in Economic Dynamics* Harvard University Press, Cambridge 1987.

- [40] G. Tauchen and R. Hussey, Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models, *Econometrica*, **59-2**, 371–396, 1991.
- [41] S. Tezuka, *Uniform random numbers: theory and practice*, Kluwer Academic, Dordrecht, 1995.
- [42] C. Thomas-Agnan, Computing a family of reproducing kernels for statistical applications, *Numerical Algorithms*, **13**, 21-32, 1996.
- [43] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.
- [44] J. F. Traub, and A. G. Werschulz, *Complexity and Information*, Cambridge University Press, Cambridge, 1998.
- [45] J. F. Traub, and H. Woźniakowski, The Monte Carlo algorithm with a pseudorandom generator, *Math. Comp.*, **58**(197), 323–339, 1992.
- [46] G. Wahba, Spline Models for Observational Data, SIAM-NSF Regional Conference Series in Appl. Math, **59**, SIAM, Philadelphia, 1990.
- [47] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor product algorithms for linear multivariate problems, *Journal of Complexity* **15-3** 402–447, 1999.
- [48] A. G. Werschulz, *The computational complexity of differential and integral equations*, Oxford University Press, New York, 1991.
- [49] H. Woźniakowski, Average case complexity of multivariate integration, *Bull. of AMS (new series)*, **24**, 185-191.
- [50] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity*, **10**, 96-128, 1994.
- [51] H. Woźniakowski, Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals, submitted for publication, 1998.

**Author’s Address:**

John Rust, Department of Economics, Yale University, P.O. Box 208264,

New Haven, CT 06520-8264, USA, email: [jrust@gemini.econ.yale.edu](mailto: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](mailto:traub@cs.columbia.edu),  
<http://www.cs.columbia.edu/~traub>,  
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](mailto:henryk@cs.columbia.edu),  
<http://www.cs.columbia.edu/~henryk>.