Annals of Operations Research

Approximating convex envelopes using linear programming

--Manuscript Draft--

Manuscript Number:	ANOR-D-18-01280
Full Title:	Approximating convex envelopes using linear programming
Article Type:	Original Research
Keywords:	convex envelope; controlled optimal stopping; non-convex optimization; linear programming; constraint sampling; function approximation
Corresponding Author:	Neeraja Sahasrabudhe, Ph.D Indian Institute of Technology Bombay Mumbai, Maharashtra INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Indian Institute of Technology Bombay
Corresponding Author's Secondary Institution:	
First Author:	Akhil Shetty
First Author Secondary Information:	
Order of Authors:	Akhil Shetty
	Saideep Reddy
	Vivek Borkar
	Neeraja Sahasrabudhe
Order of Authors Secondary Information:	
Funding Information:	Department of Science and Technology, Ministry of Science and Technology — Prof. Vivek Borkar
Abstract:	Using Oberman's characterization of the convex envelope of a function as the value function of a certain controlled optimal stopping problem, we develop a linear program for approximating the convex envelope, incorporating the ideas of linear function approximation and constraint sampling of de Farias and Van Roy to reduce the curse of dimensionality.

Noname manuscript No.
(will be inserted by the editor)

Approximating convex envelopes using linear programming

Convex Envelope via LP

Akhil Shetty[†] · **Saideep Reddy**[⋆] · **Vivek Borkar** · **Neeraja Sahasrabudhe**[⋆⋆]

Received: date / Accepted: date

Abstract Using Oberman's characterization of the convex envelope of a function as the value function of a certain controlled optimal stopping problem, we develop a linear program for approximating the convex envelope, incorporating the ideas of linear function approximation and constraint sampling of de Farias and Van Roy to reduce the curse of dimensionality.

Keywords Non-convex Optimization · Convex Envelope · Controlled Optimal Stopping · Linear Programming · Function Approximation · Constraint Sampling

1 Introduction

Non-convex optimization has become a major concern in recent years in machine learning problems. While the classical deterministic schemes yield only a local minimum, possibly a very poor one at that, the stochastic global optimiazation schemes such as simulated annealing can be very slow. It is, however, known in several situations that a clever choice of initialization of an algorithm can lead to significant improvements. One possible choice of a good starting point would be the minimum or approximate minimum of an approximate convex envelope of the function, if one such could be obtained easily. Note that the

⋆ Now with Opera Solutions, NOIDA, UP, India.

⋆⋆ Now at Indian Institute of Science Education and Research, Mohali, India.

† Now at University of California, Berkeley.

Indian Institute of Technology, Bombay
Powai, Mumbai 400076, India.
Fax: (+91 22) 2572 3707
Emails: akh9066@gmail.com, saideepreddy91@gmail.com, borkar.vs@gmail.com, neeraja.budhey@gmail.com

approximation can be allowed to be fairly inaccurate as long as its minimum is easy to find and is close to the minimum of the original function. In particular, it may not be convex but still unimodal, which is good enough for our purposes. In fact it need not even be unimodal as long as it has few local minima which can easily be worked around by simple tweaks such as multi-start. Our objective here is to provide a quick and dirty scheme for constructing an approximate convex envelope. The scheme is based on a remarkable observation of Oberman that the convex envelope is in fact the value function of an optimal stopping problem for a controlled diffusion [4]. This is a continuous time stochastic control problem and Oberman goes on to analyze the associated Hamilton-Jacobi-Bellman equation and its numerical solution [4], [6], [5]. We take a different route here, approximating the continuous time and state problem first by an optimal stopping problem for a controlled Markov chain, discrete in both time and state space. The value function satisfying the corresponding Bellman (i.e., dynamic programming) equation can be obtained by solving a linear program that goes back to Dynkin [2]. The passage from a controlled diffusion to a finite state controlled Markov chain (see, e.g., [3]) involves truncation and discretization of the state space. This can lead to a very large state space growing exponentially with the dimension of the ambient euclidean space in which the diffusion lives. This motivates a further two step approximation: the first, borrowed from [7], is to replace the value function by a linear combination of a moderate number of *fixed* basis functions (or 'features' in machine learning parlance), so that the weights thereof, not too many in number, become the new variables of optimization. This still leaves the problem of having to contend with far too many constraints. For this we adapt the idea of constraint sampling from [8] to use only a judiciously chosen subsample of constraints. It should be noted at the outset that while we borrow this idea from [8], the sampling scheme we use is tailored to suit our problem and differs from that of *ibid.* which allowed the authors to derive nice error bounds.

The details of the foregoing follow, in the following order. The next section states the problem and its equivalent controlled optimal stopping formulation due to Oberman. Section 3 describes our approximation scheme, leading to a linear programming problem via a Markov chain approximation. Section 4 details our numerical experiments on several standard test functions for non-convex optimization. Section 5 provides some pointers to future possibilities in this direction.

2 Problem Description

Let $g : \mathcal{R}^d \mapsto \mathcal{R}$ be a non-convex function. The objective is to find a function $u :=$ the convex envelope of the function g. A convex envelope is defined as:

$$u(x) = \sup\{\ v(x) \mid v \text{ convex and } v(y) \leq g(y) \text{ for all } y \in \mathcal{R}^d\}.$$

In [4], Oberman established that u can be obtained by solving the partial differential equation:

$$\max\{u(x) - g(x), -\lambda_1[u](x)\} = 0 \tag{1}$$

where $\lambda_1[u](x)$ is the smallest eigenvalue of the Hessian $D^2u(x)$. Recall that Hessian of a twice differentiable convex function is positive semidefinite. In [4], an if and only if condition of convexity of the function u is obtained in terms of viscosity solution of $-\lambda_1[u] \leq 0$ and it is proved that the convex envelope of a function g is a viscosity solution of (1). Oberman [4] arrives at this p.d.e. as the Hamilton-Jacobi-Bellman equation for a stochastic control problem that we describe next.

Consider the diffusion

$$\begin{cases} dx(t) = \sqrt{2}\kappa(t)dw(t), \\ x(0) = x_0 \end{cases}$$

where w is the one-dimensional Brownian motion and $\kappa(\cdot)$ is a measurable map $[0, \infty) \mapsto$ the unit sphere $S^n \subset \mathcal{R}^n$, satisfying the non-anticipativity condition: For $t > s \geq 0$, $w(t) - w(s)$ is independent of $\mathcal{F}_t :=$ the completion of the σ-field $\cap_{s'>s}\sigma(w(y), \kappa(y), y \leq s')$. We call such $\kappa(\cdot)$ as 'admissible' controls. The aim is to minimize the expected cost

$$J(x, \kappa(\cdot), \tau) = E^x[g(x_\tau)]$$

over all $\{\mathcal{F}_t\}$-stopping times τ and admissible $\kappa(\cdot)$. The latter makes it a 'controlled' optimal stopping problem. The convex envelope is then given by

$$u(x) = \min_{\kappa(\cdot),\tau} J(x, \kappa(\cdot), \tau)$$

which is the unique convex viscosity solution to (1) [4].

3 The linear programming formulation

We now develop a finite state approximation of the above stochastic control problem in the spirit of [3]. For this purpose, we consider a finite discretization of $\mathcal{R}^d$ given by:

$$S = \{x = [n_1\delta, n_2\delta, \ldots, n_d\delta] : -M \leq n_i \leq M\}$$

for a fixed small $\delta > 0$ and $M >> 1$. Consider the 'control space' given by

$$\mathcal{Z} = \{z = [z_1, \ldots, z_d] : z_i \in \{0, 1\} \; \forall \; 1 \leq i \leq d, \; \sum_i z_i = 1\}.$$

This replaces the unit d-sphere in which $\kappa(\cdot)$ lives. Define a controlled Markov chain $X(n), n \geq 0$, with state space S and transition probabilities:

$$p(y|x,z) = \begin{cases} 1/2 & \text{if } y_i = x_i + \delta z_i \ \forall i \\ 1/2 & \text{if } y_i = x_i - \delta z_i \ \forall i \\ 0 & \text{otherwise} \end{cases}$$

with the forbidden transition replaced by a self-loop at boundary points. This will be a discrete approximation to the controlled diffusion $x(\cdot)$. The associated dynamic programming equation is [10]

$$u(x) = \min(g(x), \min_{z \in \mathcal{Z}} \sum_y p(y|x,z)u(y))$$

with the optimal stopping time being given by

$$\tau^* := \min\{n \geq 0 : g(X(n)) = \min_{z \in \mathcal{Z}} \sum_y p(y|X(n),z)u(y)\}$$

($= \infty$ if the set on the right hand side is empty). The optimal control choices $Z(n) \in \mathcal{Z},\ 0 \leq n < \tau^*$ are given by

$$Z(n) = argmin\Big(\sum_y p(y|X(n),\cdot)u(y)\Big),$$

any tie being resolved arbitrarily. Following [2], we reformulate the problem as the linear program

$$\begin{aligned} \text{maximize} \quad & c^T u \\ \text{subject to} \quad & u(x) \leq g(x) \ \ \forall\, x \\ & u(x) \leq \sum_y p(y|x,z)u(y) \ \ \forall\, x, z, \end{aligned} \tag{2}$$

where c is any non-positive, non-zero d-vector. (There is a small difference with the framework of [2]: therein a pure optimal stopping problem is considered without the additional $\mathcal{Z}$-valued control process, ours is a *controlled* optimal stopping problem. The modification in the LP formulation this entails goes along standard lines.) A normalized c would have the interpretation of the initial distribution. Recall that from dynamic programming, we know that the optimal policy specifies the decision at a state as a function of the state alone, in particular being independent of the initial distribution. Therefore the specific choice of c is immaterial. We are, however, interested in approximating this linear program, so we take $c = [1, 1, \cdots, 1]^T$, the d-vector of all 1's, in order to avoid any bias in approximation due to the initial condition.

Clearly the state space is large and solving this LP using standard algorithms does not scale well with dimension. We propose a method based on [7], where the cost-to-go function $u : S \mapsto \mathcal{R}$ is approximated by a parametrized class of functions $S \times \mathcal{R}^K \mapsto \mathcal{R}$ parametrized by a K-dimensional parameter for a moderately large K, and then compute the optimal parameter $r^* \in \mathcal{R}^K$ that gives the best approximation for the solution u^* of the LP in (2). Specifically, we use a linear function approximation, i.e., we use a family of candidate approximations for value function that is a linear combination of K fixed basis functions. Thus consider a set of basis functions $\{\phi\}_{i=1}^K$ and define the matrix

$$\Phi = \begin{bmatrix} \vdots & & \vdots \\ \phi_1 & \dots & \phi_K \\ \vdots & & \vdots \end{bmatrix}$$

The approximate Bellman's equation for the problem in this framework is given by

$$\sum_{k=1}^{K} r_k \phi_k(x) = \min(g(x),\ \min_{z \in \mathcal{Z}} \sum_y p(y|x,z) \sum_{k=1}^{K} r_k \phi_k(y)).$$

We approximate the equivalent LP instead. The idea is solve the LP in (2) for Φr where $r \in \mathcal{R}^K$ and compute the weight vector r^* with the idea that Φr^* is an approximation for u^*. In other words our LP reduces to:

$$\begin{aligned} &\text{maximize} && c^T \Phi r \\ &\text{subject to} && \sum_y p(y|x,z) \sum_{k=1}^{K} r_k \phi_k(y) \geq \sum_{k=1}^{K} r_k \phi_k(x)\ \forall\ x, z, \\ & && g(x) \geq \sum_{k=1}^{K} r_k \phi_k(x)\ \forall\ x, \end{aligned} \tag{3}$$

where c is as before. The solution to the above problem gives an optimal weight vector r^*.

Remark 1 Clearly, the error $\|u^* - \Phi r^*\|_\infty$ depends on the choice of the basis functions for a given problem. For our numerical experiments, two types of basis functions were considered: polar and cartesian, as described later. If for a given non-convex function, certain properties such as symmetry etc. are known in advance, then one can make an informed choice of basis functions to reduces the approximation error.

Observe that there is one constraint for each state and another for each state-action pair. Thus a finer discretization of the state space blows up the number of constraints, in fact exponentially so in the dimension. One can cut down this number by sampling only the "most relevant" constraints. In [8], the authors propose a probability measure (provided the optimal policy is known) that can be used for sampling constraints. Then the solution of the resulting LP is close to that of the original LP in a precise sense. While this is an important theoretical step in the direction of tackling large number of constraints, the fact that the optimal policy has to be known a priori makes it impractical in the current scenario.

For our problem, the constraints can be divided into two types:

- **Convexity constraints:** $u(x_i) \leq \sum_j p(x_j|x_i, z)u(x_j)$.
- **Envelope constraints:** $u(x_i) \leq g(x_i)$.

We retain the entire set of envelope constraints while sampling uniformly from the set of convexity constraints, based on the observation that the former are individually crucial for our objective of obtaining a convex envelope (this is also borne by numerical experiments not reported here). On retaining them and sampling from the convexity constraints, we lose the overall convexity of the envelope in many examples, but still get a good approximation thereof with the minimum of the approximate envelope close to the minimum of the original function.

4 Numerical experiments

We consider only two dimensional problems here for ease of visualization and comparison, taken from a standard library of test functions for such purposes. We use two different sets of basis functions, inspired by the successful REctified Linear Units (RELU) in neural network literature, for simulation purposes.

- **Polar:** We consider truncated 1-dimensional linear functions and rotate them over a range of the angle θ to obtain 2-dimensional functions where required. They are defined as follows:

$$\phi_p^{i,j}(r,\theta) = \begin{cases} r - \frac{(i-1)}{2L}R & \text{if } i \text{ odd }, i \leq 2L, j \leq M, \frac{2\pi(j-1)}{M} \leq \theta \leq \frac{2\pi j}{M} \\ \frac{(2L-i+2)}{2L}R - r & \text{if } i \text{ even }, i \leq 2L, j \leq M, \frac{2\pi(j-1)}{M} \leq \theta \leq \frac{2\pi j}{M} \\ 0 & \text{otherwise} \end{cases}$$

  where, $r \in [0, R]$ for some $R > 0$. In the numerical experiments, we take $L = 20$ and $M = 20$. This gives 800 different polar basis functions.
- **Cartesian:** We consider products of 1-dimensional functions of individual variables. For example $\phi_c^{i,j}(x,y) = f^i(x)g^j(y)$ or $\phi_c^{i,j}(x,y) = h^k(\sqrt{x^2+y^2})$

where:

$$f^i(x) = \begin{cases} x - \frac{(i-1)}{2P}X & \text{if } i \text{ odd }, i \leq 2P \\ \frac{(2P-i+1)}{2P}X - x & \text{if } i \text{ even }, i \leq 2P \\ 0 & \text{otherwise} \end{cases}$$

$$g^j(y) = \begin{cases} y - \frac{(j-1)}{2Q}Y & \text{if } j \text{ odd }, j \leq 2Q \\ \frac{(2Q-j+1)}{2Q}Y - y & \text{if } j \text{ even }, j \leq 2Q \\ 0 & \text{otherwise} \end{cases}$$

$$h^k(x) = \begin{cases} r - \frac{(k-1)}{2K}R & \text{if } k \text{ odd }, k \leq 2K \\ \frac{(2K-k+1)}{2K}R - r & \text{if } k \text{ even }, k \leq 2K \\ 0 & \text{otherwise} \end{cases}$$

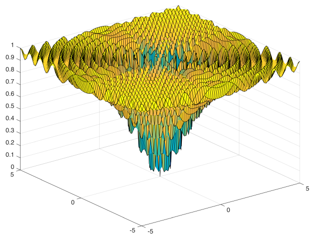
where $x \in [0, X], y \in [0, Y], r \in [0, R]$ for some $X, Y, R > 0$. For numerical experiments, we consider $P = 15, Q = 15$ and $K = 20$. That gives 940 cartesian basis functions.

As noted before, we can relax the convexity constraints but need all the envelope constraints. We sample convexity constraints uniformly with probability p. For the numerical experiments in this section, we sample each of the convexity constraints with probability $p \in \{0.05, 0.1, 0.25, 0.5\}$. Simulations were run for various one or two dimensional multimodal functions from a standard optimization function set compiled by Surjanovic and Bingham [11]. An 81×81 grid of points is taken for each function and the function values are fed to the algorithm. Constants are added when required in order to ensure that the functions considered are non-negative.

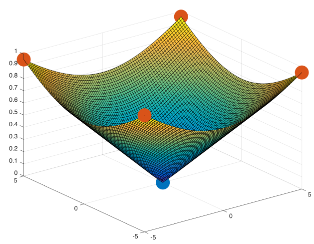
As observed, choosing cartesian coordinates instead of polar coordinates turns out to be a better strategy insofar as the minimum of the convex envelope obtained by using cartesian basis functions is closer to the actual minimum than that obtained via polar coordinates. Moreover, while sampling constraints, the envelope obtained by approximation using cartesian basis functions remains largely unchanged with changing of the sampling parameter p. This is not the case for polar basis functions. This is possibly an artefact of our rectangular domains. We shall comment more on this in the last section. If, however, we have additional information such as spherical symmetry etc. about the original function, these approximations can be improved by incorporating spherically symmetric basis functions. The numerical experiments also confirm that for smaller p, while the convergence is fast and the minimum of the approximate envelope approximates the true minimum quite well, the convexity of the envelope is lost. As p increases, the algorithm takes more time owing to a larger number of constraints, but is closer to being convex.

4.1 Examples

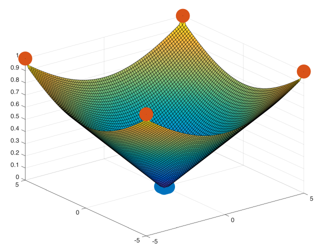
In this section, we illustrate some of the observations made above via experiments on standard test functions. Some additional interesting function-specific observations are made below for various functions. We first consider the Dropwave function. Figure 1 (a) shows the original function and the convex envelopes obtained by considering cartesian as well as polar coordinates. Figure 1 (d) illustrates the convex envelope obtained by using a mixture of cartesian and polar coordinates. Figures 2 and 3 show the convex envelope obtained after constraint sampling for different values of p.



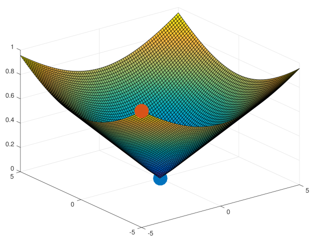
(a) Original function



(b) Envelope using Polar basis functions

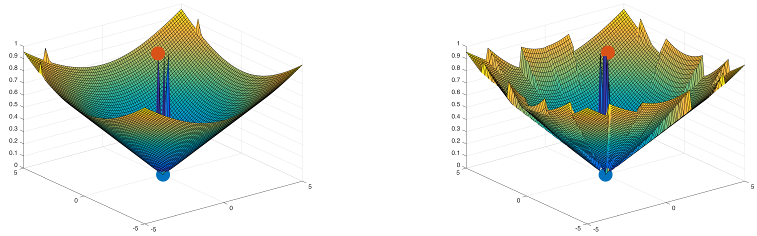


(c) Envelope using Cartesian basis function



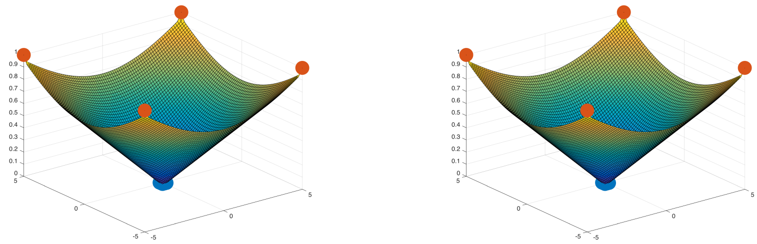
(d) Envelope using a mixture of Cartesian and Polar basis functions

Fig. 1: Convex Envelopes for Dropwave function obtained using basis function approximation.



(a) p = 0.5 (b) p = 0.05

Fig. 2: Convex Envelope for Dropwave function using polar basis functions and constraint sampling with $p = 0.5$ and $p = 0.05$.

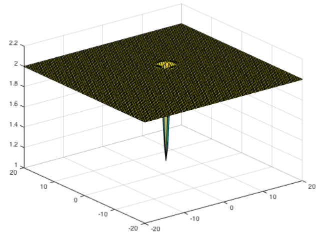


(a) p = 0.5 (b) p = 0.05

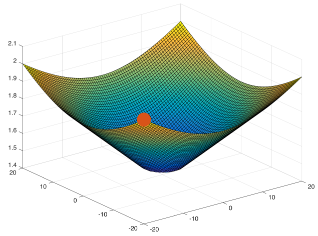
Fig. 3: Convex Envelope for Dropwave function using cartesian basis functions and constraint sampling with $p = 0.5$ and $p = 0.05$.

Both polar and cartesian basis functions yield a good approximation of the convex envelope. For both choices of basis functions, minimum is achieved at $(0, 0)$ which is the location of the global minimum of the Dropwave function. With decreasing p, the resultant envelope gets worse in case of polar basis whereas the envelope does not change much in case of cartesian basis.

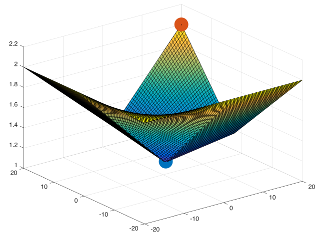
The next example in Figure 4 gives a clearer picture of how cartesian basis functions perform better than polar functions. The Easom function has a very sharp global minimum at $(-\pi, -\pi)$ which is far from the center of the domain $(0, 0)$. Note that the closest point on the grid is $(3, 3)$ and since we are using the values of the function on the grid to obtain the convex envelope, this is the best point of minimum any envelope can give. Use of polar basis functions gives an envelope with minimum at $(1, 1)$, whereas cartesian basis functions give an envelope with minimum at $(3, 3)$. A mixture of cartesian and polar basis functions with the same parameters as above gives an envelope with minimum at $(2, 2)$. Changes in envelope with changing p (Figures 5 and 6) are similar to those observed above.



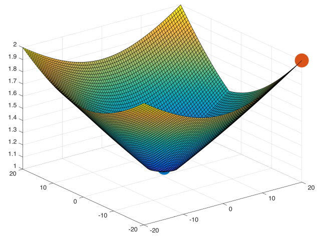
(a) Original function



(b) Envelope using Polar basis functions

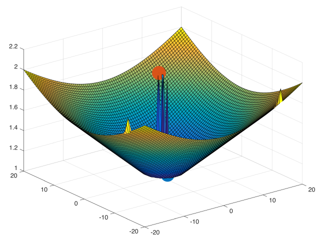


(c) Envelope using Cartesian basis functions

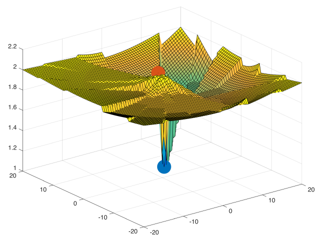


(d) Envelope using a mixture of Polar and Cartesian basis FA

Fig. 4: Easom function and convex envelopes obtained using basis function approximation

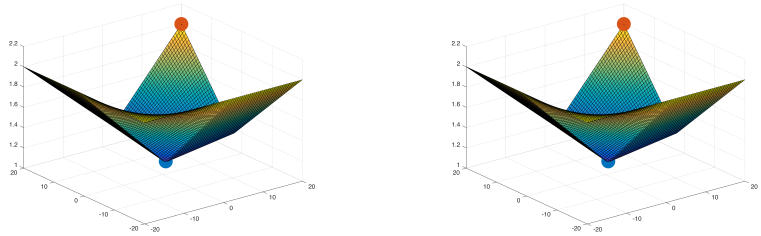


(a) p = 0.5



(b) p = 0.05

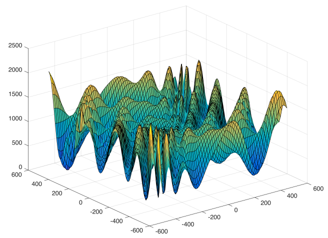
Fig. 5: Convex envelope for Easom function using polar basis functions with constraint sampling for $p = 0.5$ and $p = 0.05$.



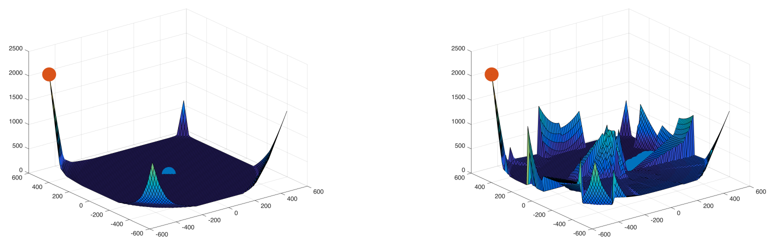
(a) p = 0.5 (b) p = 0.05

Fig. 6: Convex envelope for Easom function using cartesian basis functions with constraint sampling for $p = 0.5$ and $p = 0.05$.

Similarly, for Eggholder function (See Figure 7 (a)), approximation using cartesian basis functions performs better than the mixtures of cartesian and polar and polar basis functions. Also, in case of cartesian basis functions, reducing p brings the global minimum of the envelope closer to the global minimum of the original function without any significant loss in convexity. This is illustrated in Figure 7 and Figure 8 respectively.

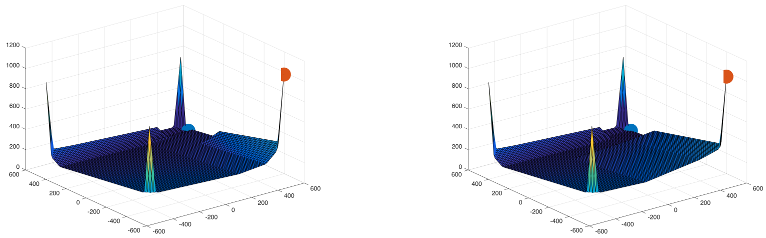


(a) Original function



(b) p = 0.5 (c) p = 0.05

Fig. 7: Convex envelope for Eggholder function using polar basis function approximations with constraint sampling for $p = 0.5$ and $p = 0.05$.

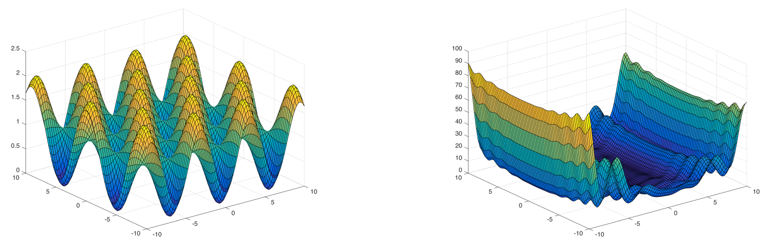


(a) p = 0.5 (b) p = 0.05

Fig. 8: Convex envelope for Eggholder function using cartesian basis function approximations respectively with constraint sampling for $p = 0.5$ and $p = 0.05$.

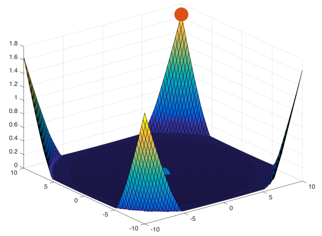
Similar behaviour was observed for Michalewicz function, Rastrigin function and Schwefel function.

For both Griewank and Levy functions (See Figure 9), polar basis functions give better results in terms of minimum of the envelope being close to the original minimum, although in the former case the improvement is marginal. In both cases, the envelopes (with both polar and cartesian basis functions) are fairly robust with changes in the constraint sampling parameter p. As observed before (for example for Eggholder function), in case of Levy function, reducing p, i.e., sampling fewer convexity constraints, improves the position of the minimum without much loss in convexity. See Figure 10 and Figure 11 for details.

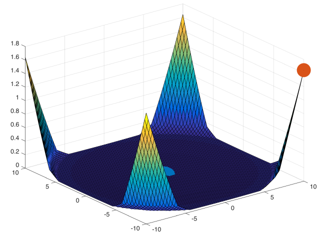


(a) Griewank function (b) Levy function

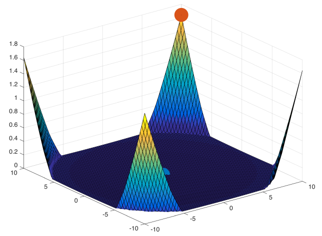
Fig. 9: Original functions



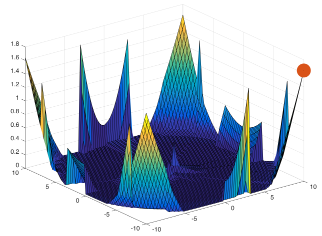
(a) Envelope using Polar basis functions



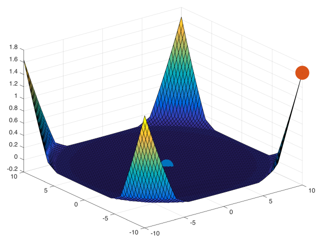
(b) Envelope using Cartesian basis functions



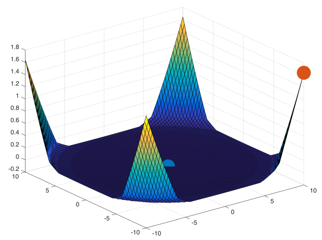
(c) p = 0.5



(d) p = 0.05

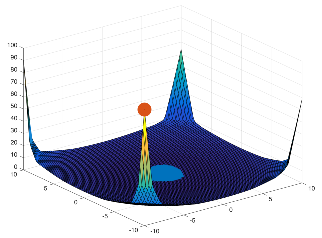


(e) p = 0.5

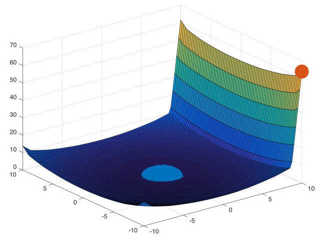


(f) p = 0.05

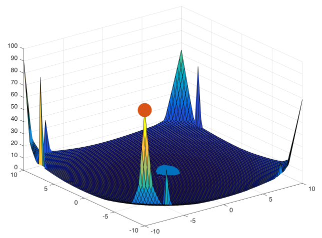
Fig. 10: Convex envelope for Griewank function using polar basis functions and cartesian basis function approximations respectively with constraint sampling for $p = 0.5$ and $p = 0.05$.



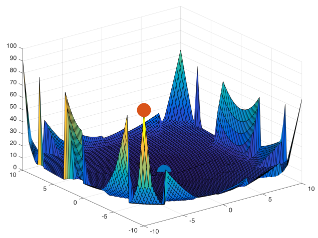
(a) Envelope using Polar basis functions



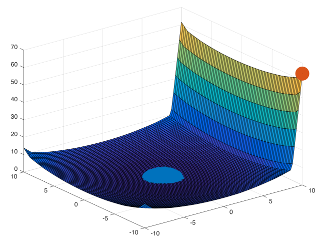
(b) Envelope using Cartesian basis functions



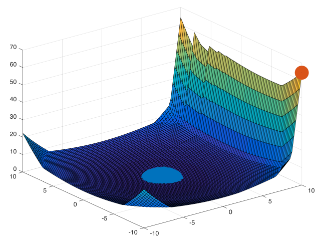
(c) p = 0.5, Polar basis function approximation



(d) p = 0.05, Polar basis function approximation



(e) p = 0.5, Cartesian basis function approximation



(f) p = 0.05, Cartesian basis function approximation

Fig. 11: Convex envelope for Levy function using polar basis functions and cartesian basis function approximations respectively with constraint sampling for $p = 0.5$ and $p = 0.05$.

In many cases instead of finding a unique point of minimum the convex envelope has multiple minima is a neighbourhood around the point closest to the original point of minimum (For example, see Figure 11).

As mentioned before, the numerical experiments were carried out for various functions and for various values of p. Given below in Table 1 and Table 2 is the tabulated observations for various cases. The minimum and minimizer for the original function, the corresponding minimum and minimizer for envelopes obtained by various choices of basis functions, along with the constraint sampling parameter p, are listed. The cases where the minimizer of the envelope is a neighbourhood around the point closest to the original minimizer are listed in the format (a to b, c to d).

Function	Basis	ProbConstr	Argmin	Min Value	Actual Argmin	Actual Min
Dropwave	polar	1	(0,0)	0	(0,0)	0
		0.5	(0,0)	0	(0,0)	0
		0.25	(0,0)	0	(0,0)	0
		0.1	(0,0)	0	(0,0)	0
		0.05	(0,0)	0	(0,0)	0
	cartesian	1	(0,0)	0	(0,0)	0
		0.5	(0,0)	0	(0,0)	0
		0.25	(0,0)	0	(0,0)	0
		0.1	(0,0)	0	(0,0)	0
		0.05	(0,0)	0	(0,0)	0
Easom	polar	1	(1,1)	1.048	(3,3)	1.0584
		0.5	(1,1)	1.048	(3,3)	1.0584
		0.25	(1,1)	1.048	(3,3)	1.0584
		0.1	(1,1)	1.048	(3,3)	1.0584
		0.05	(1,1)	1.048	(3,3)	1.0584
	cartesian	1	(3,3)	1.0584	(3,3)	1.0584
		0.5	(3,3)	1.0584	(3,3)	1.0584
		0.25	(3,3)	1.0584	(3,3)	1.0584
		0.1	(3,3)	1.0584	(3,3)	1.0584
		0.05	(3,3)	1.0584	(3,3)	1.0584
	Mixture	1	(2,2)	1.0242	(3,3)	1.0584
Eggholder	polar	1	(25.6, 25.6)	72.04	(512, 404.23)	72.9
		0.5	(25.6, 25.6)	72.66	(512, 404.23)	72.9
		0.25	(0,76.8)	72.89	(512, 404.23)	72.9
		0.1	(-140.6, -25.6)	61	(512, 404.23)	72.9
		0.05	(-140.6, -25.6)	61	(512, 404.23)	72.9
	cartesian	1	(486.4, 435.2)	71.14	(512, 404.23)	72.9
		0.5	(486.4, 435.2)	71.14	(512, 404.23)	72.9
		0.25	(486.4, 435.2)	71.14	(512, 404.23)	72.9
		0.1	(512, 435.2)	71.11	(512, 404.23)	72.9
		0.05	(512, 435.2)	71.11	(512, 404.23)	72.9
	Mixture	1	(512, 512)	65.25	(512, 404.23)	72.9
Griewank	polar	1	(0,0)	0	(0,0)	0
		0.5	(0,0)	0	(0,0)	0
		0.25	(0,0)	0	(0,0)	0
		0.1	(0,0)	0	(0,0)	0
		0.05	(0,0)	0	(0,0)	0
	cartesian	1	(-0.25 to 0.25, -0.25 to 0.25)	0	(0,0)	0
		0.5	(-0.25, 0.25)	0	(0,0)	0
		0.25	(-0.25, 0.25)	0	(0,0)	0
		0.1	(-0.25, 0.25)	0	(0,0)	0
		0.05	(-0.25, 0.25)	0	(0,0)	0

Table 1: Experimental results

Function	Basis	ProbConstr	Argmin	Min Value	Actual Argmin	Actual Min
Levy	polar	1	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
		0.5	(0.5 to 1.25, 0.5 to 1)	0	(1,1)	0
		0.25	(1,1)	0	(1,1)	0
		0.1	(1,1)	0	(1,1)	0
		0.05	(1,1)	0	(1,1)	0
	cartesian	1	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
		0.5	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
		0.25	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
		0.1	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
		0.05	(-1.25 to 1.25, -1.25 to 1.25)	0	(1,1)	0
Michalewicz	polar	1	(2,1.9)	3.17	(2.15, 1.6)	3.2
		0.5	(1.8,1.8)	3.17	(2.15, 1.6)	3.2
		0.25	(2.3,1.95)	3.17	(2.15, 1.6)	3.2
		0.1	(2.15,1.75)	3.17	(2.15, 1.6)	3.2
		0.05	(2.15, 1.6)	3.17	(2.15, 1.6)	3.2
	cartesian	1	(2.15, 1.6)	3.15	(2.15, 1.6)	3.2
		0.5	(2.15, 1.6)	3.15	(2.15, 1.6)	3.2
		0.25	(2.15, 1.6)	3.15	(2.15, 1.6)	3.2
		0.1	(2.15, 1.6)	3.15	(2.15, 1.6)	3.2
		0.05	(2.15, 1.6)	3.15	(2.15, 1.6)	3.2
Rastrigin	polar	1	(0,0)	0	(0,0)	0
		0.5	(0,0)	0	(0,0)	0
		0.25	(0,0)	0	(0,0)	0
		0.1	(0,0)	0	(0,0)	0
		0.05	(0,0)	0	(0,0)	0
	cartesian	1	(0,0)	0	(0,0)	0
		0.5	(0,0)	0	(0,0)	0
		0.25	(0,0)	0	(0,0)	0
		0.1	(0,0)	0	(0,0)	0
		0.05	(0,0)	0	(0,0)	0
Schwefel	polar	1	(-12.5,12.5)	4.09	(420,420)	4.10
		0.5	(-250,250)	4.10	(420, 420)	4.10
		0.25	(50,50)	4.10	(420, 420)	4.10
		0.1	(-50 to -12.5, -162.5 to -112.5)	0	(420, 420)	4.10
		0.05	(-150,150)	0	(420,420)	4.10
	cartesian	1	(425,425)	4.10	(420,420)	4.10
		0.5	(425,425)	4.10	(420,420)	4.10
		0.25	(425,425)	4.10	(420,420)	4.10
		0.1	(425,425)	4.10	(420,420)	4.10
		0.05	(425,425)	4.10	(420,420)	4.10
	Mixture	1	(450, 450)	3.80	(420, 420)	4.10

Table 2: Experimental results

5 Future directions

Our contributions here have been more at the level of ‘proof of concept’. There are several possible directions for further investigation.

1. The error analysis of [7] does not carry over to our case as the concerned dynamic programming operator in our case is not a contraction. A good error analysis will be extremely helpful.
2. It is worth investigating other discretization schemes for the domain. We did do some experiments with the ‘stencils’ proposed in [5] with similar results, but a systematic study with accompanying error analysis is warranted.
3. The superior performance of cartesian basis functions also appears to be at least partly an artefact of the domains we used. While rectangular domains were the de facto choice for the test function library used, a systematic study of domain effects may throw up some interesting insights.
4. A very promising research direction is to use random basis functions and invoke random projection theory to obtain probabilistic guarantees.

5. A very recent contribution by Oberman [1] looks at quasiconvex envelopes, opening up further possibilities.

Acknowledgements The research of VSB was supported in part by the project "*Approximation of High Dimensional Optimization and Control Problems*" from the Department of Science and Technology, Government of India. He thanks Prof. K. S. Mallikarjuna Rao for bringing Oberman's work to his notice and for helpful discussions.

References

1. B. Abbasi and A. M. Oberman, A partial differential equation for the strictly quasiconvex envelope, arXiv:1612.06813v1, 20 Dec., 2016.
2. E. B. Dynkin, The optimum choice of the instant of stopping a Markov process, Dokl Acad Nauk SSSR 150:238-240 (in Russian; English translation in Sov Math Dokl 4 627-629) (1963).
3. H. J. Kushner and P. G. Dupuis, Numerical Methods for Stochastic Control Problems in Continuous Time, 2nd ed., Springer Verlag, New York, 2000.
4. A. M. Oberman, The convex envelope is the solution of a nonlinear obstacle problem, Proc. Amer. Math. Soc. 135(6) 16891694 (2007).
5. A. M. Oberman, Computing the convex envelope using a nonlinear partial differential equation, Math. Models Methods Appl. Sci., 18(5) 759780 (2008).
6. A. M. Oberman and L. Silvestre, The Dirichlet problem for the convex envelope, Trans. Amer. Math. Soc. 363(11) 5871-5886 (2011).
7. D. P. de Farias and B. Van Roy, The linear programming approach to approximate dynamic programming. Oper. Res. 51(6) 850865 (2003).
8. D. P. de Farias and B. Van Roy, On constraint sampling in the linear programming approach to approximate dynamic programming, Mathematics of Operations Research, 29(3) 462-478 (2004).
9. A. Griewank and P. J. Rabier, On the smoothness of convex envelopes, Trans. Amer. Math. Soc., 322(2) 691709 (1990).
10. A. N. Shiryayev, *Optimal Stopping Rules*, Springer Verlag, New York, 1978.
11. S. Surjanovic and D. Bingham, Virtual Library of Simulation Experiments: Test Functions and Datasets (2013) Retrieved from http://www.sfu.ca/ ssurjano.