

J. Guérin · P. Marcotte · G. Savard

An optimal adaptive algorithm for the approximation of concave functions*

Received: February 16, 2001 / Accepted: December 16, 2003
Published online: December 2, 2005 – © Springer-Verlag 2005

Abstract. Motivated by the study of parametric convex programs, we consider approximation of concave functions by piecewise affine functions. Using dynamic programming, we derive a procedure for selecting the knots at which an oracle provides the function value and one supergradient. The procedure is adaptive in that the choice of a knot is dependent on the choice of the previous knots. It is also optimal in that the approximation error, in the integral sense, is minimized in the worst case.

Key words. Dynamic programming – Approximation – Adaptive algorithm

1. Introduction

The present work is motivated by a bicriterion network equilibrium problem modeled as a variational inequality (see Marcotte and Zhu [5]). In the linearization algorithm whose implementation is discussed in Marcotte, Nguyen and Tanguay [6], parametric shortest path problems have to be solved repeatedly. Since this is computationally costly, it is natural to consider the approximation of the value function of this parametric program by a piecewise linear function involving a small number of evaluation points (knots). In order to be consistent with the stopping criterion used in the linearization algorithm, the quality of the approximation has to be measured in the integral sense, i.e., with respect to the L_1 norm. This yields the problem of selecting the knots such as to minimize the approximation error, in the worst case.

In a general setting, consider a proper, concave function f defined over the interval $[0, 1]$, normalized such that $f(0) = 0$ and $f(1) = 1$. At each point $\bar{x} \in (0, 1)$ an oracle provides the value $f(\bar{x})$ and that of one supergradient $\xi \in \partial f(\bar{x})$, i.e., a point satisfying the inequality

$$f(x) \leq f(\bar{x}) + \xi(x - \bar{x}) \quad \forall x \in [0, 1].$$

J. Guérin: Département de mathématiques et génie industriel, École Polytechnique Montréal, C.P. 6079, succursale Centre-ville Montréal (Québec) H3C 3A7, Canada

P. Marcotte: CRT and DIRO, Université de Montréal, CP 6128, succursale Centre-Ville, Montréal, Canada H3C 3J7

G. Savard: GERAD and Département de mathématiques et génie industriel, École Polytechnique Montréal, C.P. 6079, succursale Centre-ville, Montréal (Québec) H3C 3A7 Canada

Mathematics Subject Classification (2000): 20E28, 20G40, 20C20

* This work was partially supported by NSERC (Canada) and FCAR (Québec).

In the sequel we denote an arbitrary supergradient ξ by $f'(\bar{x})$, even when f is not differentiable at \bar{x} . From this information we derive the under and over-approximations of f over the interval $(0,1)$:

$$L(t) = \min \left\{ \frac{f(\bar{x})}{\bar{x}}t, \frac{1-f(\bar{x})}{1-\bar{x}}(t-\bar{x}) + f(\bar{x}) \right\}$$

$$U(t) = f(\bar{x}) + f'(\bar{x})(t-\bar{x})$$

which yield the error bound $\int_0^1 (U(t) - L(t)) dt$. The aim of the present paper is to minimize the above error term through a sequential procedure for knot selection, and prove that this procedure is optimal in the sense that it minimizes the error term in the worst case. The procedure is adaptive in the sense that the selection of the current knot is dependent on the locations of previous knots.

Novak [7] and Sonnevend [9] have shown that, for the problem of approximating the integral of a convex function using function values and derivatives, the worst-case behavior of adaptive algorithms is not superior to that of passive algorithms, i.e., procedures that set the locations of all knots *simultaneously*. However, since the worst case is unlikely to occur in practice, an adaptive algorithm may take advantage of available information to produce an improved approximation. This led Sukharev [10] to the definition of *sequentially optimal* algorithms, i.e., adaptive algorithms that make optimal use, at each step, of available information. In view of the high computational complexity of such methods, we limit our analysis to algorithms where knot ordering is fixed a priori, and we have chosen the left-to-right order for simplicity. Our algorithm has the same low complexity as Sonnevend's algorithm, namely $\mathcal{O}(n)$, where n denotes the number of evaluation points. Note that the results of Novak and Sonnevend imply that all orderings are equivalent, in the worst case.

Approximation algorithms based on the bounding functions L and U have been studied in the literature under the name of "sandwich algorithms", the difference between L and U being measured with respect to the uniform, L_1 or Hausdorff distance (the Hausdorff distance between the graphs of the functions L and U). At a given iteration of a sandwich algorithm, a knot that lies in the interval of largest estimated error is determined. Fruhwirth, Burkard and Rote [3] propose, in the case of the Hausdorff distance, three subdivision rules that achieve the optimal asymptotic bound $\mathcal{O}(n^{-2})$, while a bound of the same order was also obtained by Burkard, Hamacher and Rote [2] for the uniform norm. In the paper by Rote [8] four subdivision rules are studied both from a theoretical and numerical point of view. However, two of the subdivision methods require, at each iteration, the solution of an optimization problem involving the function f itself; this violates a condition of our problem which states that no more than n function evaluations must be performed. Actually, Yang and Goh [11] showed that, if f is easy to compute, the sandwich algorithm can dispense altogether with first-order (derivative or supergradient) information.

In discussing optimal sandwich methods, Rote mentions the problem of determining an evaluation strategy that minimizes the maximal error. Our analysis brings a partial answer to this problem and improves upon previous works in two important respects:

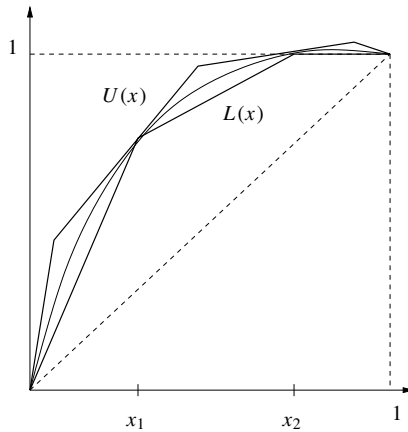


Fig. 1. Upper and lower approximation of the function f ($n = 2$)

- We obtain both the optimal convergence rate and optimal selection rules for each n .
- Our result is parameter-free: no a priori information about the function to be approximated is required.

2. Problem definition

Let f be a proper concave function defined over the unit interval $[x_0, x_{n+1}] = [0, 1]$ and normalized so that $f(0) = 0$ and $f(1) = 1$. We wish to sequentially select n points x_1, \dots, x_n in order to minimize the measure

$$\mathcal{E}(x_1, \dots, x_n) = \frac{1}{2} \int_0^1 [U(t) - L(t)] dt \tag{1}$$

where (see Figure 1)

$$L(t) = \min_{i=1, \dots, n+1} \left\{ f(x_{\sigma(i-1)}) \frac{(t - x_{\sigma(i)})}{x_{\sigma(i-1)} - x_{\sigma(i)}} + f(x_{\sigma(i)}) \frac{(t - x_{\sigma(i-1)})}{x_{\sigma(i)} - x_{\sigma(i-1)}} \right\}$$

$$U(t) = \min_{i=0, \dots, n+1} \{ f(x_i) + f'(x_i)(t - x_i) \}$$

and σ is the permutation that reorders the knots and the two endpoints from left to right:

$$0 = x_{\sigma(0)} < x_{\sigma(1)} < \dots < x_{\sigma(n)} < x_{\sigma(n+1)} = 1.$$

For a given class of functions $F \subset C[0, 1]$, let \mathcal{E}_n denote the worst-case error corresponding to an optimal selection of n knots ($n > 0$). More precisely, let A be the class of all algorithms that construct the approximation $\frac{U+L}{2}$ using the information $(f(x_1), f'(x_1), \dots, f(x_n), f'(x_n))$ obtained by evaluating the function $f \in F$ and one supergradient at n points x_1, \dots, x_n of $[0, 1]$. We consider here deterministic adaptive algorithms, where the choice of the point x_i may depend on the previous information $x_1, f(x_1), f'(x_1), \dots, x_{i-1}, f(x_{i-1}), f'(x_{i-1})$.

We denote by $\alpha \in A$ the application that maps the function $f \in F$ into its approximation $\alpha(f)$, and we define the optimal worst-case error with respect to the L_1 norm as

$$\mathcal{E}_n(a, b) = \inf_{\alpha \in A} \sup_{f \in F} \| f - \alpha(f) \|_1.$$

It is not difficult to show that the supremum in the above expression is given by the integral on the right-hand side of (1).

In the terminology of Sukharev [10], our procedure is an adaptive algorithm of the form $\alpha = (N, \phi)$, where the *information operator* is

$$N(f) = (f(x_1), f'(x_1), \dots, f(x_n), f'(x_n))$$

and the *terminal operation* ϕ is defined by

$$\phi(N(f)) = \frac{U + L}{2}.$$

It can be shown that ϕ is *central* and thus optimal in the sense that for each f , and $N(f)$ already computed, it minimizes

$$\sup_{\tilde{f} \in F_f} \| \tilde{f} - \alpha(f) \|_1,$$

where F_f is the subset of functions \tilde{f} in F that satisfy $\tilde{f}(x_i) = f(x_i)$ and $\tilde{f}'(x_i) = f'(x_i)$ for all i .

The information operator N described above is consistent with the information available and the choice of the particular terminal operation ϕ is justified by its optimality. Therefore, the construction of an approximation algorithm reduces here to the choice of the n evaluation points x_1, \dots, x_n . The optimal worst-case error \mathcal{E}_n and the optimal evaluation points can be computed through the recursion

$$\begin{aligned} \mathcal{E}_{n-k}(z^k) &= \min_{x_{k+1} \in (0,1)} \max_{(f(x_{k+1}), f'(x_{k+1})) \in C} \mathcal{E}_{n-(k+1)}(z^{k+1}) \\ &k = 0, \dots, n - 1 \end{aligned}$$

where the current information is aggregated into the vector

$$z^k = (x_1, \dots, x_k, f(x_1), \dots, f(x_k), f'(x_1), \dots, f'(x_k))$$

and C represents the set of constraints that must be satisfied by the values of $f(x_{k+1})$ and $f'(x_{k+1})$ in order to be compatible with the first k functional and supergradient values of the concave function f .

As mentioned previously, the above system is too complex to be solved in closed form and, for this reason, we limit our analysis to the identity permutation, i.e., the knots will be determined in a left-to-right fashion.

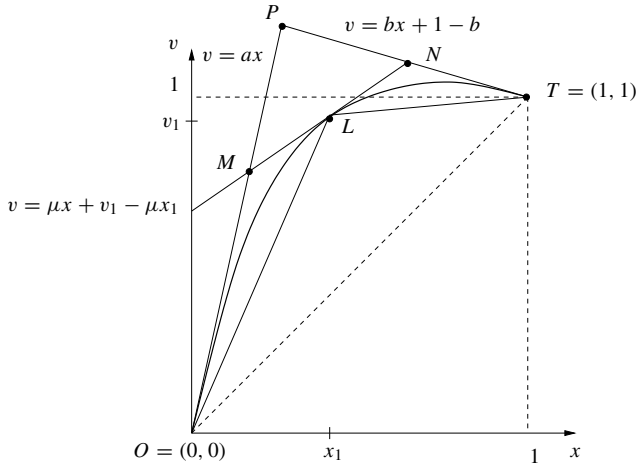


Fig. 2. Case $n = 1$

Set $a = f'(0), b = f'(1)$ and denote by $\mathcal{E}_n(a, b)$ the optimal worst-case error when knots are selected from left to right. By definition $\mathcal{E}_0(a, b)$ is equal to the area of the triangle OPT in Figure 2, i.e.

$$\mathcal{E}_0(a, b) = \frac{1}{2} \frac{(1 - b)(a - 1)}{a - b}.$$

In the case where $n = 1$, let x denote the evaluation point and set $v = f(x), \mu = f'(x)$ ¹. Since the graph of f is entirely contained within the triangle OPT , the following requirements must be met by v and μ :

$$\begin{aligned} x \leq v \leq ax & \quad \text{if} \quad x \in [0, \frac{1-b}{a-b}] \\ x \leq v \leq bx + 1 - b & \quad \text{if} \quad x \in [\frac{1-b}{a-b}, 1] \\ \frac{1-v}{1-x} \leq \mu \leq \frac{v}{x}. \end{aligned} \tag{2}$$

The error bound $\mathcal{E}_1(a, b)$ corresponds to the sum of the areas of the triangles OML and LNT of Figure 2 and can be expressed in term of \mathcal{E}_0 :

$$\mathcal{E}_1(a, b) = \min_{x \in (0,1)} \max_v \max_{\mu} \left\{ xv\mathcal{E}_0\left(\frac{x}{v}a, \frac{x}{v}\mu\right) + (1-x)(1-v)\mathcal{E}_0\left(\frac{1-x}{1-v}\mu, \frac{1-x}{1-v}b\right) \right\},$$

where v and μ must satisfy the geometric constraints (2), and the scaling factors multiplying \mathcal{E}_0, a, b and μ are derived from elementary geometric arguments. Now, for an arbitrary number n , the worst-case error term can be defined recursively as

¹ From now on, we drop knot indices, with the exception of Figure 2, where they have been retained in order to avoid a collision of the symbol x with itself.

$$\mathcal{E}_n(a, b) = \min_{x \in (0,1)} \max_v \max_{\mu} \left\{ xv\mathcal{E}_0\left(\frac{x}{v}a, \frac{x}{v}\mu\right) + (1-x)(1-v)\mathcal{E}_{n-1}\left(\frac{1-x}{1-v}\mu, \frac{1-x}{1-v}b\right) \right\} \tag{3}$$

under constraints (2). Any minimizer x of $\mathcal{E}_n(a, b)$ is called *optimal*. The next point of evaluation is then set to a minimizer of $\mathcal{E}_{n-1}\left(\frac{1-x}{1-f(x)}f'(x), \frac{1-x}{1-f(x)}b\right)$, and so on to the n th knot. Our main result follows.

Theorem 1. *The optimal worst-case error is equal to*

$$\mathcal{E}_n(a, b) = \frac{1}{2(n+1)^2} \frac{(a-1)(1-b)}{(a-b)} \tag{4}$$

and the minimum in (3) is achieved at the point

$$x^* = \frac{1}{(n+1)^2} \left(1 + 2n \frac{1-b}{a-b}\right). \tag{5}$$

Note that, whenever f is concave increasing and no a priori information is available on the slopes a and b , i.e., $a = +\infty$ and $b = 0$, the above formula collapses to

$$\mathcal{E}_n(a, b) = \frac{1}{2(n+1)^2}.$$

3. Proof of the theorem

The proof of Theorem 1 is lengthy, due to the many cases and subcases that have to be probed, and we only provide an outline. The reader interested in the complete proof is referred to Guérin [4]. The proof proceeds by induction on n . The result clearly holds for $n = 0$. For $n \geq 1$ we evaluate the expression

$$\mathcal{R}_n(x) = \max_v \max_{\mu} \left\{ xv\mathcal{E}_0\left(\frac{x}{v}a, \frac{x}{v}\mu\right) + (1-x)(1-v)\mathcal{E}_{n-1}\left(\frac{1-x}{1-v}\mu, \frac{1-x}{1-v}b\right) \right\},$$

working backwards with respect to the two “max” operators. For fixed x and v , let us consider the function

$$\phi(\mu) = 2 \left[xv\mathcal{E}_0\left(\frac{x}{v}a, \frac{x}{v}\mu\right) + (1-x)(1-v)\mathcal{E}_{n-1}\left(\frac{1-x}{1-v}\mu, \frac{1-x}{1-v}b\right) \right].$$

Using the induction hypothesis to eliminate the terms \mathcal{E}_0 and \mathcal{E}_{n-1} , ϕ can be expressed as

$$\phi(\mu) = A \frac{v - \mu x}{a - \mu} + B \frac{(1-x)\mu - (1-v)}{n^2(\mu - b)},$$

where $A = ax - v$ and $B = 1 - v - (1-x)b$ are nonnegative scalars.

The rational function ϕ , which is illustrated on Figure 3, possesses two local optima, denoted μ^- and μ^+ , respectively. Its maximum μ_{\max} on $[(1-v)/(1-x), v/x]$ is

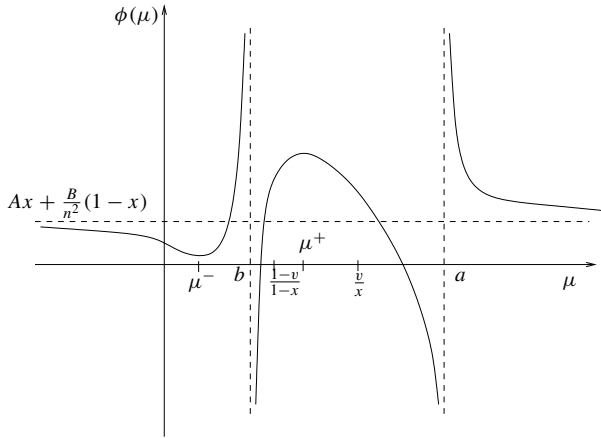


Fig. 3. The function ϕ (case $\mu_{\max} = \mu^+$)

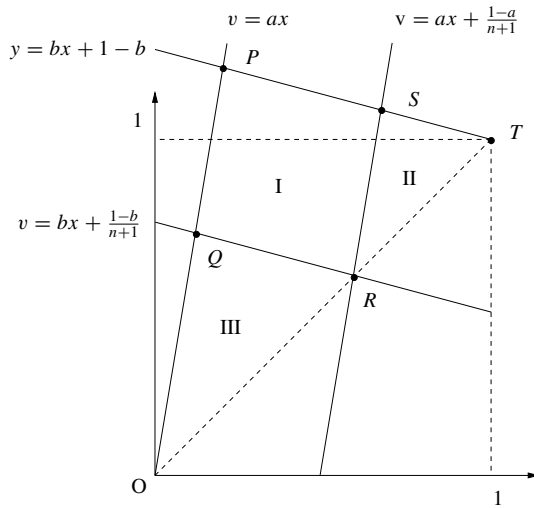


Fig. 4. Three cases for μ

achieved either at μ^+ or at one of the endpoints of the interval. This yields three cases, each one corresponding to an area of the triangle $O P T$ in Figure 4. The three areas I, II and III are defined, respectively, as the quadrilateral $P Q R S$, the triangle $R S T$ and the triangle $O R Q$. The maximum of ϕ occurs at μ^+ if (x, v) belongs to area I, at $(1 - v)/(1 - x)$ if (x, v) belongs to area II and at v/x if (x, v) belongs to area III.

For notational purposes we introduce, for fixed x , the scaling:

$$\psi(v) = n^2(a - b)\phi(\mu_{\max}).$$

The value v_{\max} at which ψ reaches its maximum defines a piecewise smooth function of x consisting of three linear and one quadratic pieces. The analysis must consider two cases, depending whether n is larger or less than $(a - 1)/(1 - b)$. The function v_{\max} ,

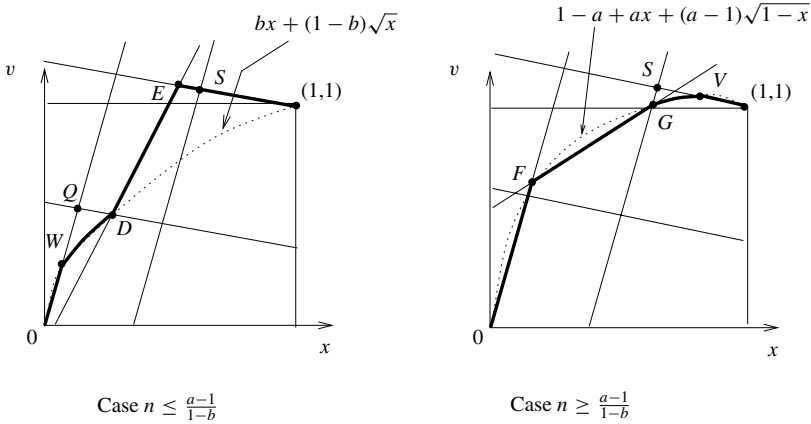


Fig. 5. The function v_{\max}

illustrated on Figure 5, is defined as

$$v_{\max} = \begin{cases} ax & \text{if } x \in (0, W_1] \\ bx + (1 - b)\sqrt{x} & \text{if } x \in [W_1, D_1] \\ \frac{a + b}{2}x + \frac{2(n + 1) - a - (2n + 1)b}{2(n + 1)^2} & \text{if } x \in [D_1, E_1] \\ bx + (1 - b) & \text{if } x \in [E_1, 1) \end{cases}$$

in the case $n \leq (a - 1)/(1 - b)$, and by

$$v_{\max} = \begin{cases} ax & \text{if } x \in (0, F_1] \\ \frac{a + b}{2}x + \frac{2(n + 1) - a - (2n + 1)b}{2(n + 1)^2} & \text{if } x \in [F_1, G_1] \\ 1 - a + ax + (a - 1)\sqrt{1 - x} & \text{if } x \in [G_1, V_1] \\ bx + (1 - b) & \text{if } x \in [V_1, 1) \end{cases}$$

if $n \geq (a - 1)/(1 - b)$. (Subscript “1” refers to the x -coordinate of a point.)

Next, we determine the minimum of the function $R_n(x)$ over the interval $(0, 1)$ by computing the minimal value of R_n over each subinterval. Next, we check that the minimum occurs at the point x^* , with minimal value given by the formula of Theorem 1. The minimum occurs when (x, v) belongs to area I and

$$v_{\max} = \frac{a + b}{2}x + \frac{2(n + 1) - a - (2n + 1)b}{2(n + 1)^2}.$$

This brings the argument to its conclusion.

4. Numerical tests

Based on the optimal formula provided by Theorem 1, we implemented the adaptive algorithm DYN, which computes the optimal location of n points from left to right or

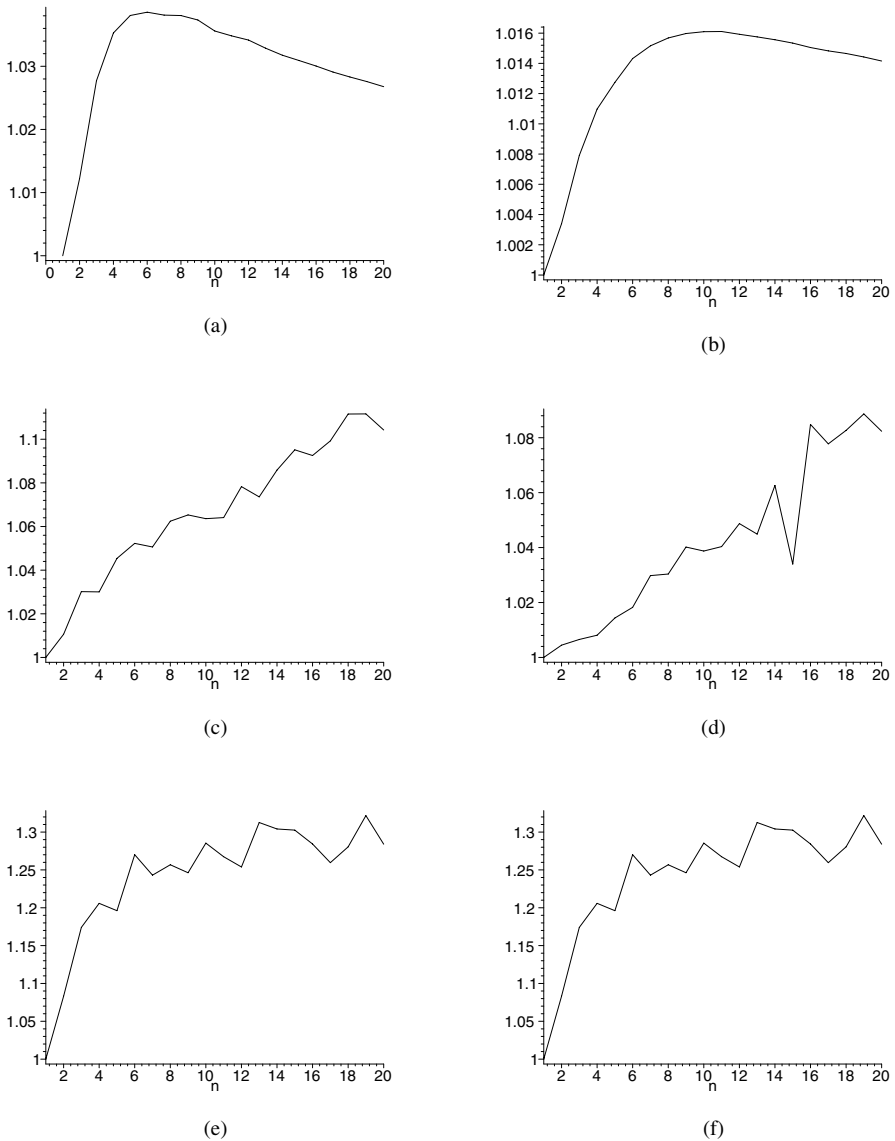


Fig. 6. Ratio SONN over DYN of average errors for (a) smooth concave increasing, (b) smooth concave, (c) PL concave increasing, (d) PL concave, (e) PS concave increasing and (f) PS concave functions

right to left. The choice of the direction is determined by a heuristic procedure based on a priori information. The performance of DYN is compared with that of SONN, the optimal passive algorithm proposed by Sonnevend [9]. The computational complexity of both algorithms is $\mathcal{O}(n)$ function and derivative evaluations.

The performance of DYN and SONN was tested on sets of randomly generated concave functions. Three functional forms were considered: smooth, piecewise linear (PL) and piecewise smooth (PS). For each form, two samples were produced: one consisting of concave increasing functions and the other of general concave functions.

For each sample, the average error was computed for values of n ranging from 1 to 20. This is consistent with the range considered in the bicriteria traffic equilibrium problem discussed in the introduction. The results from both algorithms, which were compared by taking the ratio of the average error of SONN over the average error of DYN, are illustrated in Figures 6(a)–6(f).

- On average, DYN performed much better than SONN on all six samples. Gains in accuracy were largest for piecewise smooth functions and least for smooth functions, with gains for piecewise linear functions falling in between.
- For nearly all functions tested, DYN performed at least as well as SONN. In the case where DYN's performance is worse, the difference in accuracy is at most 2%.
- On some specific functions, the gain in accuracy achieved by DYN is as high as 400%. Large gains were observed on functions exhibiting strong curvature near the endpoints.

References

1. Bellman, R., Dreyfus, S.: Applied Dynamic Programming. Princeton University Press, Princeton, 1962
2. Burkard, R.E., Hamacher, H.W., Rote, G.: Sandwich approximation of univariate convex functions with an application to separable convex programming. *Naval Res. Logistics* **38**, 911–924 (1991)
3. Fruhwirth, B., Burkard, R.E., Rote, G.: Approximation of convex curves with application to the bicriterial minimum cost flow problem. *Eur. J. Oper. Res.* **42**, 326–338 (1989)
4. Guérin, J.: Une méthode adaptative pour l'approximation de fonctions concaves croissantes. Mémoire de maîtrise de l'École Polytechnique de Montréal, 2000
5. Marcotte, P., Zhu, D.L.: Equilibria with infinitely many differentiated classes of customers. *Complementary and Variational Problems, State of the Art*. Jong-Shi Pang and Michael Ferris, (eds.), SIAM, Philadelphia, 1997, pp. 234–258
6. Marcotte, P., Nguyen, S., Tanguay, K.: Implementation of an efficient algorithm for the multiclass traffic assignment problem. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Jean-Baptiste Lesort, (ed.), Pergamon 1996, pp. 217–236
7. Novak, E.: Quadrature formulas for convex classes of functions. *International Series of Numerical Mathematics* **112** Birkhäuser Verlag, Basel, 1993, pp. 283–296
8. Rote, G.: The convergence rate of the sandwich algorithm for approximating convex functions. *Computing* **48**, 337–361 (1992)
9. Sonnevend, G.: Optimal passive and sequential algorithms for the approximation of convex functions in $L_p[0, 1]^s$, $p = 1, \infty$. *Constructive function theory '81*. Sofia, 1983
10. Sukharev, A.G.: Minimax models in the theory of numerical methods. *Theory and Decision Library Series B Vol. 21*. Kluwer Academic Publishers, 1992
11. Yang, X.Q., Goh, C.J.: A method for convex curve approximation. *Eur. J. Oper. Res.* **97**, 205–212 (1997)