Control and Cybernetics

vol. **29** (2000) No. 2

On the choice of statistical model for one-dimensional P-algorithms*

by

James M. Calvin[1] **and Antanas Žilinskas**[2]

[1]Department of Computer and Information Science
New Jersey Institute of Technology
Newark, NJ 07102-1982, USA

[2]Institute of Mathematics and Informatics, Akademijos str. 4
Vilnius, LT2600, Lithuania

Abstract: Algorithms based on statistical models compete favorably with other global optimization algorithms as proved by extensive testing results. Recently, techniques were developed for theoretically estimating the rate of convergence of global optimization algorithms with respect to the underlying statistical models. In the present paper these techniques are extended for theoretical investigation of P-algorithms without respect to a statistical model. Theoretical estimates may eliminate the need for lengthy experimental investigation which previously was the only method for comparison of the algorithms. The results obtained give new insight into the role of the underlying statistical model with respect to the asymptotic properties of the algorithm which will be useful for the implementation of new versions of the algorithms.

Keywords: optimization, statistical models, convergence.

1. Introduction

Statistical models are useful for modeling the global behavior of a complicated multimodal function and the normal uncertainty of the researcher with respect to the features of the function, Törn and Žilinskas (1989). However, the choice of a concrete model may not be obvious, and may be guided by conflicting goals. For example, the Wiener process seems well-grounded as a global model of complicated one-dimensional multimodal functions. The Wiener model has been the

*This project has been funded in part by the National Research Council of the USA under

most common model used for the implementation of global optimization algorithms (Boender and Romeijn, 1995; Žilinskas, 1981; Calvin, 1999; Kushner, 1962, 1964; Locatelli, 1997; Ritter, 1990; Groch, Vidigal, and Director, 1985; Locatelli and Schoen, 1995), although the local features of non-differentiable Wiener process paths differ essentially from the local features of smooth objective functions. For a long time there has been interest in implementations based on the smooth function statistical models, though progress has been blocked by the computational complexity of such implementations. The problem of local inadequacy was resolved by using a dual *Wiener process/quadratic function* model, Törn and Žilinskas (1989); Žilinskas (1981). From a practical point of view the P*-algorithm based on such a dual model is sufficiently efficient as shown by the results of extensive testing in Törn and Žilinskas (1989); Žilinskas (1981). However, a theoretically justified algorithm based on a statistical model of smooth functions had not been known until recently.

Extension of the approach to statistical models of smooth objective functions was initiated in Calvin and Žilinskas (1999), where an approximation method was proposed enabling the construction of a P-algorithm whose computational complexity is similar to that of the algorithm based on the Wiener model. The availability of two competing models suggests their comparison. Theoretical estimates of the convergence rate of P-algorithms with respect to the underlying statistical models are formulated in probabilistic terms in Calvin (1999); Calvin and Žilinskas (1999), but the applicability to a specific practical case is not always obvious.

In the present paper the convergence rate of P-algorithms based on Wiener and smooth function models is compared for neutral conditions, i.e. for general assumptions on smoothness of an objective function without respect to a statistical model. Under both algorithms, the order of convergence is the same up to a constant. However, the constant depends on features of the objective function and a choice of the method parameter ϵ. This result gives new insight into the role of the underlying model in determining the asymptotic properties of the algorithm. Until recently, the only way to asses the convergence rate and to compare efficiency of statistical model-based methods was implementation and experimental testing. The new techniques of theoretical investigation may be used prior to experimentation to prevent the time consuming implementation of theoretically unsatisfactory versions.

2. P-algorithm

In this section we will describe the P-algorithm, which is motivated by probabilistic considerations. The reader should keep in mind that the probabilistic considerations serve only to motivate the algorithm; we will investigate the convergence properties of the algorithm in the following section.

the case where f does not satisfy stronger regularity conditions, such as convexity, unimodality, or availability of a Lipschitz constant. Let the minimal value $M = \min_{0\le x\le 1} f(x)$ be attained at the point x^*, and assume that $F(x) > M$ for $x \neq x^*$. The stochastic process $\{\xi(x) : 0 \le x \le 1\}$ is accepted as a statistical model of the objective function; parameters of the model can be estimated from an initial sampling of function values at points uniformly distributed in $[0, 1]$. Fix $\epsilon > 0$. The n-th observation of the function value is performed by the P-algorithm at the point

$$x_{n+1} = \arg\max_{0\le x\le 1} P\{\xi(x) < M_n - \epsilon | \xi(x_i) = y_i, i = 1, \ldots, n\}, \tag{2.1}$$

where $x_i, y_i = f(x_i)$ are the results of observations at previous minimization steps and $M_n = \min_{1\le i\le n} y_i$. Let us denote the ordered observation points by

$$0 = x_0^n < x_1^n < \cdots < x_n^n = 1,$$

and the corresponding function values by $y_i^n = f(x_i^n)$, $i \le n$. If $\xi(x)$ is a Markov process, then the conditional distribution of the value $\xi(x)$ in (2.1) would depend only on neighboring function values and the maximization in (2.1) might be replaced by the following simpler procedure: For each subinterval $[x_{i-1}^n, x_i^n]$, $i \le n$, calculate

$$\max_{x_{i-1}^n \le x \le x_i^n} P\{\xi(x) < M_n - \epsilon \,|\, \xi(x_{i-1}^n) = y_{i-1}^n,\ \xi(x_i^n) = y_i^n\}, \tag{2.2}$$

and for the interval with the largest probability, calculate the point that maximizes the probability in (2.2); this point is the new observation point x_{n+1}. The details of the implementation of the Wiener process version of the algorithm (2.1) may be found in Žilinskas (1981).

The Wiener process corresponds to general assumptions on uncertainty in global optimization, Törn and Žilinskas (1989), and therefore it seems an adequate model to describe the global behavior of complicated multimodal functions. However, it does not accurately represent the local features of typical smooth objective functions, since the sampling functions of the Wiener process are everywhere non-differentiable almost surely. In Calvin and Žilinskas (1999), a statistical model of a smooth objective function is justified, i.e. a stationary Gaussian stochastic process $\xi(x)$ with zero mean, unit variance and the correlation function $r(\cdot)$, which we assume to be of the form

$$r(t) = 1 - \frac{1}{2}\lambda_2 t^2 + \frac{1}{4!}\lambda_4 t^4 + o(t^4) \tag{2.3}$$

as $t \to 0$, for finite λ_2, λ_4. We further assume that

$$|d^4 r(t)/dt^4 - \lambda_4| = O(|t|),\ -r''(t) = \lambda_2 + O(|\log^{-a}|t||)$$

for some $a > 1$ as $t \to 0$, and also that $r(t)\log(t) \to 0$ as $t \to \infty$. These assump-

sample functions, Lindgren (1972). The algorithm (2.2) with underlying smooth function model is an approximated P-algorithm whose details are discussed in Calvin and Žilinskas (1999).

In the remainder of this section we introduce parameters that help characterize the convergence rate of algorithms. The magnitude of the error depends in the limit on two quantities associated with the objective function f. We can think of these as a *global* characteristic $\Gamma_\beta(\epsilon)$, depending on the parameters ϵ, β, and a *local* characteristic Λ. For $\epsilon > 0$ and $\beta > 0$, define the global characteristic by

$$\Gamma_\beta(\epsilon) \triangleq \int_{x=0}^{1} \left(1 + \frac{f(x) - M}{\epsilon}\right)^{-\beta} dx. \tag{2.4}$$

Note that $\Gamma_\beta(\epsilon)$ is decreasing in β and increasing in ϵ. The larger values of $0 < \Gamma_\beta(\epsilon) \leq 1$ correspond to more difficult minimization problems due to the concentration of function values near to the global minimum. A constant objective function has maximal index of difficulty $\Gamma_\beta(\epsilon) = 1$. The case of a constant objective function is the worst case for many global optimization algorithms, see Törn and Žilinskas (1989). The meaning of these characteristics is illustrated by some examples.

EXAMPLE 2.1 *Let* $f(x) = (x - x^*)^2$, *where* $x^* = 1/2$. *Then*

$$\Gamma_{1/2}(\epsilon) = \int_{x=-1/2}^{1/2} \left(1 + x^2/\epsilon\right)^{-1/2} dx = \sqrt{\epsilon} \ln\left(1 + \frac{1}{2\epsilon} + \frac{1}{2\epsilon}\sqrt{1 + 4\epsilon}\right)$$
$$\sim \sqrt{\epsilon} \ln(1 + 1/\epsilon)$$

as $\epsilon \downarrow 0$. *Taking* $\beta = 2$, *we have*

$$\Gamma_2(\epsilon) = \int_{x=-1/2}^{1/2} \left(1 + x^2/\epsilon\right)^{-2} dx$$
$$= \frac{1}{2\left(1 + \frac{1}{4\epsilon}\right)} + \sqrt{\epsilon} \arctan\left(\frac{1}{2\sqrt{\epsilon}}\right) \sim \frac{\pi}{2}\sqrt{\epsilon}$$

as $\epsilon \downarrow 0$.

EXAMPLE 2.2 *Let* $f(x) = |x - x^*|$, *where* $x^* = 1/2$. *Then*

$$\Gamma_{1/2}(\epsilon) = 2\int_{x=0}^{1/2} (1 + x/\epsilon)^{-1/2} dx = 4\epsilon\left(\sqrt{1 + \frac{1}{2\epsilon}} - 1\right)$$

and

$$\Gamma_2(\epsilon) = 2\int_{x=0}^{1/2} (1 + x/\epsilon)^{-2} dx = \frac{2\epsilon}{}$$

For the local characteristic, we assume that there exists a positive number α such that

$$\lim_{n\to\infty} n^{\alpha} \sup_{|x|\le 1/n} f(x^* + x) - M = \Lambda(f) > 0. \tag{2.5}$$

For example, if $f(x) = a\,|x - x^*|$, then $\alpha = 1$ and $\Lambda(f) = a$. If f is smooth with positive second derivative at x^*, then $\alpha = 2$ and $\Lambda(f) = f''(x^*)/2$.

3. Asymptotic normalized error bounds

According to Kushner (1962); Calvin and Žilinskas (1999), the P-algorithm for the Wiener and smooth function models is defined by the following procedures.

If the Wiener process model is used, for each subinterval $[x_{i-1}^n, x_i^n]$, $i \le n$, calculate

$$\rho_i^n \triangleq \frac{x_i^n - x_{i-1}^n}{(y_{i-1}^n - M_n + \epsilon)(y_i^n - M_n + \epsilon)}. \tag{3.1}$$

The next observation is made in the interval with the maximal value of ρ_i^n, at the point

$$x_{n+1} = \frac{x_{i-1}^n + x_i^n}{2} - \frac{x_i^n - x_{i-1}^n}{2} \frac{y_i^n - y_{i-1}^n}{\left(y_{i-1}^n + y_i^n - 2M_n + 2\epsilon\right)}.$$

If the smooth process model is used, for each subinterval $[x_{i-1}^n, x_i^n]$, $i \le n$, calculate

$$\gamma_i^n \triangleq \frac{x_i^n - x_{i-1}^n}{\sqrt{y_{i-1}^n - M_n + \epsilon} + \sqrt{y_i^n - M_n + \epsilon}}. \tag{3.2}$$

The new observation is made in the interval with the maximum value of γ_i^n, at the location

$$x_{n+1} = x_{i-1}^n + \frac{(x_i^n - x_{i-1}^n)(y_{i-1}^n - M_n + \epsilon)}{y_{i-1}^n - M_n + \epsilon + \sqrt{(y_{i-1}^n - M_n + \epsilon)^2 + (y_i^n - y_{i-1}^n)(y_{i-1}^n - M_n + \epsilon)}}.$$

The following theorem summarizes our results.

THEOREM 3.1 *Fix $\epsilon > 0$. Let an objective function $f(x)$ be continuous over the minimization interval and have local characteristic $\Lambda(f)$ with exponent α at the unique global minimizer x^*, and global characteristic $\Gamma_\beta(\epsilon)$. Then the P-algorithm, based either on the Wiener or on a smooth function model, converges to the global minimum in the sense that the error $\Delta_n \downarrow 0$ and in particular*

for the algorithm based on a smooth process model, and

$$\limsup_{n\to\infty} n^{\alpha}\Delta_n \le 4\Lambda(f)\Gamma_2^2(\epsilon) \tag{3.4}$$

for the algorithm based on the Wiener process.

As previously noted, $\Gamma_{1/2}(\epsilon) > \Gamma_2(\epsilon)$. However, this does not necessarily imply that the Wiener-process-based algorithm is more efficient for a smooth objective function: the reason is that a good choice of ϵ may be different for the two cases. While choosing ϵ a trade-off between its influence on the convergence rate and the global distribution of trial points should be taken into account. This question is analyzed in the next section.

We now proceed to the proof of the theorem.

Proof. We begin with the proof of (3.3). Since the observations are dense in $[0,1]$ and f is continuous, $M_n \downarrow M$ and

$$\sum_{i=1}^{n}\gamma_i^n \to \frac{1}{2}\int_{x=0}^{1}\frac{dx}{\sqrt{f(x)-M+\epsilon}} = \frac{\Gamma_{1/2}(\epsilon)}{2\sqrt{\epsilon}}. \tag{3.5}$$

Let x_L^n and x_R^n be the observation points to the left and right, respectively, of the minimizer x^*; that is, for some j_n,

$$x_L^n = x_{j_n}^n \le x^* \le x_{j_n+1}^n = x_R^n.$$

Let $y_L^n = f(x_L^n)$ and $y_R^n = f(x_R^n)$ be the corresponding function values. Let

$$\gamma_s^n \triangleq \frac{x_R^n - x_L^n}{\sqrt{y_L^n - M_n + \epsilon} + \sqrt{y_R^n - M_n + \epsilon}} \tag{3.6}$$

and

$$\gamma^n = \max_{i\le n}\gamma_i^n,\ \gamma_n = \min_{i\le n}\gamma_i^n.$$

Since the subintervals are eventually bisected, Calvin and Žilinskas (1999), $\gamma^n/\gamma_n \to 2$.

Because $y_L^n - M_n \to 0$ and $y_R^n - M_n \to 0$,

$$\frac{x_R^n - x_L^n}{2\gamma_s^n\sqrt{\epsilon}} = \frac{\sqrt{y_L^n - M_n + \epsilon} + \sqrt{y_R^n - M_n + \epsilon}}{2\sqrt{\epsilon}} \to 1, \tag{3.7}$$

which, combined with (3.5), implies that

$$\frac{n(x_R^n - x_L^n)}{\gamma_s^n/\frac{1}{n}\sum_{i=1}^{n}\gamma_i^n} \to 2\sqrt{\epsilon}\,\Gamma(\epsilon). \tag{3.8}$$

Since

$$\frac{\gamma_s^n}{\gamma_n}$$

we have that

$$\liminf_{n\to\infty} \frac{\gamma_s^n}{\frac{1}{n}\sum_{i=1}^n \gamma_i^n} \geq \liminf_{n\to\infty} \frac{\gamma_n}{\gamma^n} = \frac{1}{2}. \tag{3.9}$$

Re-writing, we obtain

$$n(x_R^n - x_L^n) = \frac{(x_R^n - x_L^n)}{2\gamma_s^n\sqrt{\epsilon}} \frac{\gamma_s^n}{\frac{1}{n}\sum_{i=1}^n \gamma_i^n} 2\sqrt{\epsilon} \sum_{i=1}^n \gamma_i^n, \tag{3.10}$$

and applying (3.7), (3.9), and (3.5) to the three terms on the right-hand side of (3.10), we conclude that

$$\limsup_{n\to\infty} n(x_R^n - x_L^n) \leq 4\sqrt{\epsilon}\frac{\Gamma_{1/2}(\epsilon)}{2\sqrt{\epsilon}} = 2\Gamma_{1/2}(\epsilon). \tag{3.11}$$

We now turn our attention from the gap surrounding the global minimizer to the error $\Delta_n = M_n - M$. Because of our assumption on the existence of $\Lambda(f)$,

$$\limsup_{n\to\infty} \frac{\Delta_n}{\Lambda(f)(x_R^n - x_L^n)^2} \leq 1.$$

Therefore, by (3.11),

$$\begin{aligned}\limsup_{n\to\infty} \frac{n^\alpha \Delta_n}{\Lambda(f)} &= \limsup_{n\to\infty} \frac{\Delta_n}{\Lambda(f)(x_R^n - x_L^n)^\alpha}[n(x_R^n - x_L^n)]^\alpha \\ &\leq 4\Gamma_{1/2}^2(\epsilon),\end{aligned} \tag{3.12}$$

which proves (3.3).

We now derive a similar upper bound for the Wiener process-based algorithm; for details see Calvin, 1999. Recall that the quantity to be maximized over all intervals (analogous to γ_i^n) is

$$\rho_i^n \triangleq \frac{x_i^n - x_{i-1}^n}{(y_{i-1}^n - M_n + \epsilon)(y_i^n - M_n + \epsilon)}.$$

Also define

$$\rho_n = \min_{i\leq n} \rho_i^n, \ \rho^n = \max_{i\leq n} \rho_i^n.$$

A similar analysis to that carried out for the smooth case yields

$$\frac{x_R^n - x_L^n}{\rho_s^n} \to \epsilon^2, \ \frac{\rho^n}{\rho_n} \to 2,$$

and

$$\sum_{i=1}^n \rho_i^n \to \int_0^1 \frac{dt}{\Gamma_2(\epsilon) - \Gamma_1(\epsilon)t} = \frac{1}{\Gamma_2(\epsilon)}\ln\left(\frac{\Gamma_2(\epsilon)}{\Gamma_2(\epsilon) - \Gamma_1(\epsilon)}\right).$$

Therefore,

$$\limsup_{n\to\infty} n(x_R^n - x_L^n) = \limsup_{n\to\infty} \frac{(x_R^n - x_L^n)}{\rho_s^n} \frac{\rho_s^n}{\frac{1}{n}\sum_{i=1}^n \rho_i^n} \sum_{i=1}^n \rho_i^n$$
$$\leq 2\Gamma_2(\epsilon), \tag{3.13}$$

and so under the Wiener process-based algorithm,

$$\limsup_{n\to\infty} \frac{n^\alpha \Delta_n}{\Lambda(f)} = \limsup_{n\to\infty} \frac{\Delta_n}{\Lambda(f)(x_R^n - x_L^n)^\alpha} [n(x_R^n - x_L^n)]^\alpha \leq 4\Gamma_2^2(\epsilon). \tag{3.14}$$

This completes the proof of (3.4). ■

The asymptotic behavior of the P-algorithm depends mainly on the properties of the objective function and not on the stochastic process used to justify the algorithm. The dependence on the stochastic model enters only in the constant factor.

Since the same form of conditional mean is used for both algorithms, the way the observations are allocated depends on the conditional variance. The algorithm based on the smooth model spreads observations more uniformly over the interval.

Examination of (2.4) shows that $\Gamma_\beta(\epsilon)$ is large when f spends a lot of time near the global minimum M. In this case, it is not possible to concentrate the search effort as much near x^* since the promising region is large. In contrast, if f has a narrow "spike" at x^*, then the algorithm can concentrate the search effort there and so the error is relatively small (in this case $\Gamma_\beta(\epsilon)$ is relatively small).

If the same value of ϵ were chosen for both versions of P-algorithm then the algorithm based on the Wiener process model would converge faster than the algorithm based on the smooth function model. Such a conclusion might seem surprising at first. However, the choice of ϵ should normally be larger for the first than for the second version of the algorithm, as shown by the analysis in next section.

Thus far we have considered only a fixed objective function f. We now briefly describe the situation when f is a random function with the same smoothness properties previously assumed. Assuming that the smooth function f has a unique global minimizer that has an absolutely continuous distribution on $[0, 1]$, we can derive descriptions of the normalized point processes of observations near x^*. Basically, x^* is asymptotically uniformly distributed in the interval formed by the two nearest observations, and the subintervals are eventually bisected. There exists a sequence of stopping times $\{n_k : k \geq 1\}$ such that the point process of observations near x^* (suitably normalized) converges to a

smoothness assumption on f at x^*, allow us to prove (Calvin, 1999; Calvin and Žilinskas, 1999) that

$$\frac{n_k^2 \Delta_{n_k}}{C} \xrightarrow{\mathcal{D}} \min\{U^2, (1-U)^2\} \tag{3.15}$$

for some normalizing random variable C (which depends on the algorithm), where U is a uniformly distributed random variable on the unit interval. Here $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution; i.e., $X_n \xrightarrow{\mathcal{D}} X$ for random variables X_n, X if $P(X_n \leq x) \to P(X \leq x)$ for all x such that $P(X = x) = 0$. For the two algorithms under consideration, we obtain

$$\frac{n_k^2 \Delta_{n_k}}{\frac{1}{2} f''(x^*) \Gamma_\beta(\epsilon)^2} \xrightarrow{\mathcal{D}} \min\{U^2, (1-U)^2\}, \tag{3.16}$$

where $\beta = 2$ for the Wiener process case and $\beta = 1/2$ for the smooth process case.

4. Asymptotic distribution of trial points

Let us denote the smallest subinterval containing point x after n minimization steps $\Delta(n, x)$.

THEOREM 4.1 *The ratio of lengths of subintervals satisfies the inequality*

$$\frac{1}{2}\left(\frac{f(x) - M + \epsilon}{f(z) - M + \epsilon}\right)^2 \leq \liminf_{n\to\infty} \frac{\Delta(n,x)}{\Delta(n,z)} \leq \limsup_{n\to\infty} \frac{\Delta(n,x)}{\Delta(n,z)}$$
$$\leq 2\left(\frac{f(x) - M + \epsilon}{f(z) - M + \epsilon}\right)^2$$

for the Wiener model-based P-algorithm and

$$\frac{1}{2}\sqrt{\frac{f(x) - M + \epsilon}{f(z) - M + \epsilon}} \leq \liminf_{n\to\infty} \frac{\Delta(n,x)}{\Delta(n,z)} \leq \limsup_{n\to\infty} \frac{\Delta(n,x)}{\Delta(n,z)}$$
$$\leq 2\sqrt{\frac{f(x) - M + \epsilon}{f(z) - M + \epsilon}}$$

for the smooth function model-based P-algorithm.

Proof. Let us consider the Wiener model case. The analysis for the smooth function statistical model case is similar. The trial points are dense everywhere in the minimization interval. For continuously differentiable functions $x_i^n - x_{i-1}^n \to 0$ implies that the point x maximizing (2.2) is equal to

$$x_i^n + x_{i-1}^n \qquad ((\quad n \quad)^2$$

i.e., the intervals are eventually bisected, Calvin and Žilinskas (1999), by both versions of P-algorithm. For sufficiently large n the criterion value ρ_i^n of subinterval (x_{i-1}^n, x_i^n) is not larger than the criterion value of any other subinterval before the subdivision, i.e. it is not larger than double the criterion value of any other subinterval. Estimating the limit of the ratios of corresponding criterion values (3.1) yields the proof of the theorem for the Wiener model case. ■

The choice of ϵ is important not only for the constant $\Gamma_\beta(\epsilon)$ in the estimate of the convergence rate but also for the global distribution of trial points. The value of ϵ should be chosen small to ensure high convergence rate, but it also should be chosen sufficiently large to ensure globality of search, i.e. sufficient density of trial points in the subintervals of "medium" function values. The ratio $1/R$ of the length of a subinterval in the vicinity of the global minimizer and of an interval with "medium" function values is implied by the choice of ϵ satisfying the inequalities

$$\frac{1}{2}\left(\frac{(F-M)/2+\epsilon}{\epsilon}\right)^2 \leq R \leq 2\left(\frac{(F-M)/2+\epsilon}{\epsilon}\right)^2$$

for the Wiener model-based algorithm, and by the choice of ϵ satisfying the inequalities

$$\frac{1}{2}\sqrt{\frac{(F-M)/2+\epsilon}{\epsilon}} \leq R \leq 2\sqrt{\frac{(F-M)/2+\epsilon}{\epsilon}}$$

for the smooth function model-based algorithm, where F denotes the maximum of $f(x)$. These inequalities suggest the choice $\epsilon = \frac{F-M}{2\sqrt{R}}$ for the first case and the choice $\epsilon = \frac{F-M}{2R^2}$ for the second case. The theoretically justified choice of ϵ corresponds to a formula empirically found in Žilinskas (1981), for the Wiener model-based P-algorithm.

Let us consider **Example 2.1**. The ratio $R = 4$ corresponds to the choice $\epsilon = 1/(8\sqrt{2})$ for the Wiener model-based algorithm and $\epsilon = 1/128$ for the smooth function-based algorithm. Since $\Gamma_2(1/8\sqrt{2})^2 = 0.2181$ and $\Gamma_{1/2}(1/128)^2 = 0.1845$, we may conclude that for a quadratic objective function the algorithm based on a smooth function model converges faster than the Wiener model-based algorithm if an appropriate choice of ϵ ensures similar global behavior for both algorithms. The objective function of **Example 2.2** is non-differentiable at the minimum point. In this case the relation between $\Gamma_2(1/(8\sqrt{2}))^2 = 0.0226$ and $\Gamma_{1/2}(1/128)^2 = 0.0487$ shows a faster convergence of the Wiener model-based algorithm. For sufficiently small ϵ the examples represent two rather broad classes of objective functions: either smooth or non-differentiable at the mini-

5. Conclusions

For general assumptions on an objective function the versions of P-algorithm based on Wiener and smooth function models have the same order of convergence. Under appropriate choice of ϵ ensuring similar global behavior of both versions of the algorithm, the constant in the estimate of the convergence rate is better for the Wiener model-based algorithm if $f(x)$ is non-differentiable at the global minimum point. Similarly, the constant is better for the smooth function model-based algorithm if $f(x)$ is differentiable at the global minimum point.

References

BOENDER, G. and ROMEIJN, E. (1995) *Stochastic methods.* In: R. Horst and P. Pardalos, eds., *Handbook of Global Optimization.* KAP, 829–869.

CALVIN, J. (1999) Convergence rate of the p-algorithm. New Jersey Institute of Technology, Computer and Information Science Report No. **99-3**.

CALVIN, J. and ŽILINSKAS, A. (1999) On convergence of the p-algorithm for one-dimensional global optimization of smooth functions. *Journal of Optimization Theory and Applications*, **102**, 3, 479–495.

GROCH, A., VIDIGAL, L. and DIRECTOR, S. (1985) A new global optimization method for electronic circuit design. *IEEE Trans. on Circuits and Systems*, **32**, 160–170.

KUSHNER, H. (1962) A versatile stochastic model of a function of unknown and time-varying form. *Journal of Mathematical Analysis and Applications*, **5**, 150–167.

KUSHNER, H. (1964) A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, **86**, 97–106.

LINDGREN, G. (1972) Local maxima of gaussian fields. *Ark. Math.*, **10**, 195–218.

LOCATELLI, M. (1997) Bayesian algorithms for one-dimensional global optimization. *Journal of Global Optimization*, **10**, 57–76.

LOCATELLI, M. and SCHOEN, F. (1995) An adaptive stochastic global optimization algorithm for one-dimensional functions. *Annals of Operations Research*, **58**, 263–278.

RITTER, K. (1990) Approximation and optimization on the wiener space. *Journal of Complexity*, **6**, 337–364,

TÖRN, A. and ŽILINSKAS, A. (1989) *Global Optimization.* Springer.

ŽILINSKAS, A. (1981) Two algorithms for one-dimensional multimodal minimization. *Mathematische Operationsforschung und Statistik, ser. Optimization*, **12**, 53–63.