SAND 2000-1395J

Evolutionary Pattern Search Algorithms for Unconstrained and Linearly Constrained Optimization

William E. Hart
Sandia National Laboratories
Optimization/Uncertainty Estimation Department
P. O. Box 5800, MS 1110
Albuquerque, NM 87185-1110
Phone: 505-844-2217 Fax: 505-845-7442
wehart@sandia.gov
http://www.cs.sandia.gov/~wehart

May 16, 2000

Abstract

We describe a convergence theory for evolutionary pattern search algorithms (EPSAs) on a broad class of unconstrained and linearly constrained problems. EPSAs adaptively modify the step size of the mutation operator in response to the success of previous optimization steps. The design of EPSAs is inspired by recent analyses of pattern search methods. Our analysis significantly extends the previous convergence theory for EPSAs. Our analysis applies to a broader class of EPSAs, and it applies to problems that are nonsmooth, have unbounded objective functions, and which are linearly constrained. Further, we describe a modest change to the algorithmic framework of EPSAs for which a non-probablistic convergence theory applies. These analyses are also noteworthy because they are considerably simpler than previous analyses of EPSAs.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

1 Introduction

Evolutionary pattern search algorithms (EPSAs) are a class of real-valued evolutionary algorithms (EAs) that can be applied to unconstrained minimization problems

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & x \in \mathbf{R}^n \end{array} \tag{1}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$, as well as bound constrained problems

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & x \in \mathbf{R}^n \\ & l_i \leq x_i \leq u_i, \quad i = 1, \ldots, n \end{array} \tag{2}$$

where $l_i, u_i \in \mathbf{R}$, and $l_i < u_i$. Like evolutionary programing (EP) [16] and evolutionary strategies (ESs) [31], EPSAs are *adaptive* EAs that modify the mutation *step length* during optimization. However, EPSAs have an absolute step length that is used to generate a finite set of offsets, while EPs and ESs generate mutations by adding a continuous random variable that is scaled by a step length parameter. Also, EPSAs use a global step length parameter instead of the per-individual step length parameter commonly used by ESs and EPs.

Hart [22] describes EPSAs and proves a weak stationary-point convergence theory for these methods for arbitrary unconstrained or bound-constrained continuously differentiable objective functions. This analysis of EPSAs can be characterized by the following properties:

- **Type of Adaptivity**: EPSAs adapt the step length dynamically, and they do not limit the range of adaptation.
- **Objective Function**: The class of objective functions to which this analysis applies includes smooth, nonconvex problems.
- **Convergence Analysis**: The convergence analysis for EPSAs guarantees convergence near a stationary point.
- **Approximations**: No approximations are used in this analysis. The mathematical formulation of EPSAs exactly captures their empirical behavior.

Although none of these features distinguishes EPSAs from previous analyses by itself, this analysis is particularly distinguished by its ability to exactly capture the analytic behavior of a class of adaptive EAs on a broad class of nonlinear problems. Thus we do not approximate the stochastic process underlying EPSAs, as is commonly done in the study of self-adaptive EP and ESs.

Additionally, this analysis is distinguished by our focus on a stationary-point convergence theory. Although stationary-point convergence is implied by some proofs of global convergence for EAs (e.g. see Rudolph [30]), these proofs limit the adaptivity of the EA (e.g. limiting the mutation step scale above a fixed threshold). Consequently, these analyses fail to describe the convergence behavior of *self-adaptive* methods for EP and ESs that have proven quite successful in practice [3, 4, 15, 16]. Furthermore, analyses of self-adaptive EAs have focused on convex, unimodal problems, and thus they provide limit insight into the convergence behavior of these EAs on the nonconvex, multimodel problems to which they are commonly applied. Our motivation for developing EPSAs is to develop a better understanding of the role of adaptivity of the mutation step length in real-coded EAs on general classes of nonconvex multimodal problems.[1]

The main result of this paper is to describe a new convergence theory for EPSAs on problems of the form

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & x \in \Omega = \{x \in \mathbf{R}^n \mid l \leq Ax \leq u\}, \end{array} \tag{3}$$

where $l, u \in \mathbf{R}^m \bigcup \{\pm\infty\}$ and $A \in \mathbf{Q}^{m \times n}$. Note that this class of problems encompasses both problems (1) and (2). We use techniques recently developed by Audet and Dennis [2] for analyzing pattern search methods to show how the convergence theory for EPSAs can be extended in several directions. The new convergence theory provides a natural notion of convergence for nonsmooth problems, and the analysis allows for points where the objective function is infinite. These results greatly extend the theoretical range of applicability for EPSAs. For example, the ability to optimize problems with infinite objective functions enables the solution of problems for which the objective function may fail to return a value at points in the feasible domain (e.g. this can happen in engineering design applications which rely on numerical solution techniques that fail to converge at certain design points [2, 13]).

Our analysis also extends the model of EPSAs in several ways. First, it allows members of the initial population to lie in $\mathbf{R}^n$. Our previous analysis imposed the somewhat artificial condition that the initial points lie in $\mathbf{Q}^n$ [22]. Second, the EPSAs that we analyze may change the set of possible mutation offsets that they use. Finally, we describe modest changes to the basic design

[1] We use the term *step length* to describe the absolute step scale for methods like EPSAs as well as scale factors for random variables that are added to perform mutation in methods like ES and EP.

of EPSAs that enable these EA's to converge deterministically even though the algorithm remains stochastic. These changes involve assumptions on the selection process and on how the mutation offsets are applied to the best point in the population.

In the next section we describe EPSAs, and we review the probabilistic weak stationary-point convergence theory in Hart [22]. The next section generalizes this probabilistic convergence theory. Finally, we describe EPSAs for which a nonprobabilistic convergence theory applies.

2 Evolutionary Pattern Search Algorithms

Consider the pseudo code in Figure 1. This is a modest generalization of the class of EPSAs considered by Hart [22]. The algorithmic structure of EPSAs is formulated so that these EAs can be cast as stochastic pattern search methods [22], a class of randomized derivative-free optimizers for which a weak stationary-point convergence theory has been proven. Although there are many possible ways to define EAs in this manner, the EPSAs defined in Figure 1 come closest to capturing the basic algorithmic framework of canonical EAs like genetic algorithms (GAs) [19], EP and ESs.

An EPSA is initialized with an initial step length Δ_0 and with points in Ω. A finite set of matrices, $\mathcal{S}$, are chosen so that for all $S \in \mathcal{S}$ the columns of S form a positive basis [9]. The column vectors in S, $\{s_1, \ldots, s_{m_S}\}$, represent the mutation offsets that are applied to a point in an EPSA's population, and the array η is used to indicate whether or not a given offset vector has been applied. An EPSA is allowed to select a new set of mutation offsets after an improving mutation is generated from the best point in the population or after all mutation offsets off of the best point, x_t^*, have been sampled. However, for bound-constrained (and linearly constrained) problems additional restrictions are placed upon the choice of S when the population approaches a constraint boundary (see Section 4).

The calls to the **selection**, **crossover** and **compose** functions are exactly the same as a generic EA (e.g. see the generic EA described by Hart [22]). The **selection** function stochastically selects a subset of points from the previous population with a bias towards more optimal points, the **crossover** operator combines two points in $\overline{X}$ to generate a new trial point, and the **compose** function combines the points from the previous population and the new points generated via crossover and mutation to form the next population. The mutation step involves the random selection of a step in S, an evaluation of whether a mutation in that direction (scaled by Δ_t) is feasible, and an update to $\hat{x}_i$ if it is feasible. This is equivalent to the types of steps made by an EP

```
Given $\Delta_0 \in \mathbf{R}^{>0}$
$\mathcal{S} = \{S_1, \dots, S_k\}$, $S_i \in \mathbf{Q}^{n \times m_i}$, where the columns of $S_i$ form a positive basis
Select an initial population $X_0 = \{x_1^0, \dots, x_N^0\}$, $x_i^0 \in \Omega$
$x_0^* = \arg\min\{f(x_1^0), \dots, f(x_N^0)\}$
Select $\overline{S} \in \mathcal{S}$; let $\eta = \{0\}^{m_{\overline{S}}}$
Repeat $t = 0, 1, \dots$
    $\overline{X} = \text{selection}(X_t)$
    For $i = 1 : N$
        If (unif() $< \chi$) then $\hat{x}_i = \text{crossover}(\overline{x}_{\text{uint}(N)}, \overline{x}_{\text{uint}(N)})$
        Else $\hat{x}_i = \overline{x}_{\text{uint}(N)}$
    For $i = 1 : N$
        If (unif() $< \mu$) then
          $j = \text{uint}(m_{\overline{S}})$
          If ($\hat{x}_i + \Delta_t \cdot \overline{s}_j$ is feasible)
            If ($\hat{x}_i == x_t^*$) $\eta_j = 1$
            $\hat{x}_i = \hat{x}_i + \Delta_t \cdot \overline{s}_j$
          Else $\eta_j = 1$
    $X_{t+1} = \text{compose}(X_t, \hat{X})$
    $x_{t+1}^* = \arg\min\{f(x_1^{t+1}), \dots, f(x_N^{t+1})\}$
    If ($f(x_{t+1}^*) < f(x_t^*)$)
        If ($\exists s \in \overline{S}$ s.t. $x_{t+1}^* = x_t^* + s$) $\Delta_{t+1} = \Delta_t * \theta_t$
        Select $\overline{S} \in \mathcal{S}$; let $\eta = \{0\}^{m_{\overline{S}}}$
    ElseIf ($|\eta| == m_{\overline{S}}$)
        Select $\overline{S} \in \mathcal{S}$; let $\eta = \{0\}^{m_{\overline{S}}}$
        $\Delta_{t+1} = \Delta_t \lambda_t$
    Else
        $\Delta_{t+1} = \Delta_t$
Until some stopping criterion is satisfied
```

Figure 1: **Pseudo Code for EPSAs**

or ES, but with mutation applied with probability μ and with the restriction that the probability distribution of possible mutation steps is finite. Finally, the function **uint**(i) generates an integer in $\{1, \ldots, i\}$ uniformly at random.

The update to the step length, Δ_t, is much like pattern search methods. If an improving point is generated in the current iteration, then Δ_t may be increased if x^*_{t+1} was generated by a mutation from x^*_t (and η is reset). If all mutation offsets of x^*_t have been examined, then Δ_t is shrunk and η is reset. Otherwise $\Delta_{t+1} = \Delta_t$.

2.1 Convergence Theory

Various conditions are placed upon EPSAs to prove our convergence results. Mild restrictions are placed upon the **selection** and **compose** functions to ensure that (a) the best point in the population is selected with probability of at least $\pi > 0$ in each iteration and (b) the best point from the previous population and the newly generated points, $X_t \bigcup \hat{X}$, is always included in X_{t+1}. The **crossover** function is also restricted to generate a point such that crossover$(x, y) \in \{x_1, y_1\} \times \{x_2, y_2\} \times \ldots \times \{x_n, y_n\}$, which is consistent with standard coordinate-wise crossover operators (e.g. two-point crossover). Further, note that other approaches can be used to apply crossover (e.g. two parents are used to generate two children), as long as the crossover method used generates points composed from the coordinate values of the parents; this excludes intermediate crossover methods used with ESs, for example. Let $\Phi = \overline{\Phi} \bigcup \underline{\Phi}$, where $|\Phi| = \nu < \infty$, $\phi \in \mathbf{Q}$ for all $\phi \in \Phi$, $\phi \geq 1$ for all $\phi \in \overline{\Phi}$ and $\phi < 1$ otherwise. We require that $\theta_t \in \overline{\Phi}$ and $\lambda_t \in \underline{\Phi}$. Thus θ_t are expansion factors for Δ_t and λ_t are contraction factors.

In additon to these restrictions, which apply to all EPSAs that we consider in this paper, our previous analysis made the following assumptions. First, S contains a single rational matrix. Further, $x^0_i \in \mathbf{Q}^n$, $i = 1, \ldots, N$, and $\Delta_0 \in \mathbf{Q}^{>0}$. Let $\tau = \tau_n/\tau_d$, $\tau_n, \tau_d \in \mathbf{N}$, and $\tau_n > \tau_d$. We restrict the contraction factor for Δ_t to $\underline{\Lambda} = \{\tau^{\kappa_0}\}$, where $\kappa_0 \in \mathbf{Z}^{<0}$, and $\overline{\Lambda} = \{1, \tau^{\kappa_1}, \ldots, \tau^{\kappa_d}\}$, where $\kappa_i \in \mathbf{Z}^{>0}$.

Given these conditions on the design of EPSAs, the main results in Hart [22] are Theorems 1 and 2, which prove a stationary-point convergence for unconstrained and bound constrained stochastic pattern search respectively. The subsequent analysis in the paper shows that EPSAs can be cast as stochastic pattern search algorithms. Let $L_\Omega(y) = \{x \in \Omega \mid f(x) \leq f(y)\}$ and

$L(y) = L_{\mathbf{R}^n}(y)$. Further, let

$$p_i(t) = \begin{cases} l_i & \text{if } t < l_i \\ t & \text{if } l_i \leq t \leq u_i \\ u_i & \text{if } t > u_i \end{cases}$$

and consider the projection of $x \in \mathbf{R}^n$ onto the feasible region of problem (2),

$$Q(x) = \sum_{i=1}^{n} p_i(x_i) e_i,$$

where e_i is the ith standard coordinate vector. Note that x is a stationary point of (2) if and only if $q(x) = Q(x - g(x)) - x = 0$, where $g(x)$ is the gradient of f at x (e.g., see [6, 8]). In the bound constrained theory, the quantity $q(x)$ plays the role of the gradient $g(x)$, providing a continuous measure of how close x is to a constrained stationary point. If $g(x) = 0$ such that the constraints are not active, then clearly $q(x) = Q(x - g(x)) - x = Q(x) - x = 0$. Otherwise, $q(x)$ is a continuous function that is zero at a point where $g(x)_i \leq 0$ if $x_i = u_i$, $g(x)_i \geq 0$ if $x_i = l_i$ and $g(x)_i = 0$ if $l_i < x_i < u_i$. Thus $q(x) = 0$ ensures that x is a constrained stationary point.

Theorem 1 ([22], Theorem 1) *Let $\Omega = \mathbf{R}^n$ and let $L(x_0)$ be compact. If $f : \mathbf{R}^n \to \mathbf{R}$ is continuously differentiable on an open neighborhood of $L(x_0)$, then for the sequence of iterates $\{x_k\}$ produced by an EPSA,*

$$P\left(\liminf_{k\to\infty} \|g(x_k)\| = 0\right) = 1.$$

Theorem 2 ([22], Theorem 2) *Let Ω be a bound-constrained feasible domain, and let $L_\Omega(x_0)$ be compact. If $f : \mathbf{R}^n \to \mathbf{R}$ is continuously differentiable on an open neighborhood of $L_\Omega(x_0)$, then for the sequence of iterates $\{x_k\}$ produced by an EPSA,*

$$P\left(\liminf_{k\to\infty} \|q(x_k)\| = 0\right) = 1.$$

This convergence guarantee is weak, since it only implies that the gradient is sampled infinitely often near a stationary point. Thus it is possible that $\limsup_{k\to\infty} \|g(x_k)\| > 0$ (e.g. see the example in Audet [1] for a simple pattern search method). However, the sequence of iterates generated by a pattern search method is monotone nonincreasing and bounded below on a compact set, so $\lim_{k\to\infty} f(x_k) = \hat{f}$ for some fixed value. Note that this is a "global" convergence analysis since it guarantees convergence to a stationary point from any starting point. This terminology is unfortunate in that convergence to a *global* minimizer of the function is *not* implied. However, "locally convergent" is reserved for another use for nonlinear optimization (e.g., see [10]).

3 Unconstrained Analysis

This section describes how our previous analysis of EPSAs can be both generalized and simplified to solve problem (3) for unconstrained problems (where A is the identity matrix, $l = \{-\infty\}^m$ and $u = \{\infty\}^m$). Our analysis is based on the recent work of Audet and Dennis [2], which reconsiders the convergence theory of generalized pattern search methods. Following Audet and Dennis, our analysis allows EPSAs to be applied to functions with weaker continuity assumptions, and they can also be applied to problems where the objective function is undefined for some points in the feasible domain.

The following analysis differs significantly from our previous analysis [22] in that we do not prove a convergence theory for stochastic pattern search methods and then explicitly show that EPSAs can be cast as stochastic pattern search. Instead, we provide a direct analysis of EPSAs. We believe that this direct analysis will more clearly illustrate the mathematical structure of EPSAs that is being exploited to provide a proof of convergence. Further, it is increasingly tedious to demonstrate the equivalence of EPSAs and pattern search methods as we consider more generalized versions of EPSAs.

3.1 Overview

Consider $\{x_t^*\}$, the sequence of the best points found so far for each iteration of an EPSA. Note that x_t^* is a member of the t-th population because of the restrictions imposed on EPSAs. Our convergence analysis makes the standard assumption that all points sampled by the EPSA lie in a compact set [2]. A reasonable sufficient condition for this to hold is that $L_{\mathbf{R}^n}(y)$ is compact. However, our analysis does not make this assumption because we allow discontinuities and even $f(x) = \infty$ for some x, so $L_{\mathbf{R}^n}(y)$ may not be closed. However, we could assume that the set is bounded or precompact [17].

Our analysis of EPSAs focuses on the convergence properties of the sequence $\{x_t^*\}$. The points generated as mutation offsets of the points in this sequence can be viewed as the core trial steps of a simple pattern search method. The values of all other points generated by the EPSA are relevant in our analysis only to the extent that they may generate a point that is the better than x_t^*; these points are akin to the *search* steps considered in the generalize pattern search method of Audet and Dennis [2].

Since $\{x_t^*\}$ lie in a compact set, there exist convergence subsequences of this sequence. We say that a convergent subsequence $\{x_k^*\}_{k\in K}$ (for some set of indeces K) is a *refining subsequence* if $\Delta_{k_i} > \Delta_{k_{i+1}}$ for all $k_i \in K$ and $\lim_{k\in K} \Delta_k = 0$. We focus on convergent subsequences for EPSAs since the conditions required to ensure that the entire sequence converges are rather restrictive (e.g. see Torczon [33]), which makes it difficult to define EPSAs that reflect the common elements of canonical EAs. The following proposition shows that there exists a refining subsequence for the sequence of best points found by an EPSA with probability one.

Proposition 1 *There exists a refining subsequence of $\{x_t^*\}$ with probability one.*

The following theorem describes limit points of refining subsequences for general nonsmooth functions. A natural generalization of the notion of a gradient for nonsmooth functions is the generalized directional derivative [7]. The generalized directional derivative of f at x in the direction s is

$$f^o(x;s) = \limsup_{y\to x, t\downarrow 0} \frac{f(y+ts) - f(y)}{t}.$$

Note that if $f^o(x;s) \geq 0$, then f is increasing in the direction s. Thus a local minimum of a nonsmooth function is defined by a point where $f^o(x;s') \geq 0$ for all $s' \in \mathbf{R}^n$ and where there exists a direction s such that $f^o(x;s) = 0$.

Theorem 3 *Let $\hat{x}$ be the limit of a refining subsequence $\{x_k^*\}_{k\in K}$. If f is Lipshitz in the neighborhood of $\hat{x}$ then there exists $S \in \mathcal{S}$ such that for all $s \in S$, $f^o(x;s) \geq 0$ if $x_k^* + \Delta_k s$ is feasible for infinitely many $k \in K$.*

For unconstrained problems, this theorem shows that there is a positive basis $S \in \mathcal{S}$ that defines directions for which the generalized directional derivatives of f are positive. This is perhaps the strongest possible result for EPSAs on general nonsmooth problems. If the refining subsequence converges to a nondifferentiable point, then it may be possible for $f^o(x;\epsilon s) \geq 0$ for every direction s in a given basis and for all $\epsilon > 0$ (e.g. see Torczon [32]). For constrained problems, this result indicates that the interesting search steps are those that *conform* to the boundary of the feasibile domain (see Section 4).

The next theorem extends the previous result when f is strictly differentiable at a limit point $\hat{x}$. A point x is strictly differentiable if $\bigtriangledown f(x)$ exists and $\bigtriangledown f(x)^T w = \lim_{y\to x, t\downarrow 0} \frac{f(y+tw)-f(y)}{t}$ for all $w \in \mathbf{R}^n$ [7].

Theorem 4 *Let $\Omega = \mathbf{R}^n$, and let $\hat{x}$ be the limit of a refining subsequence. If f is Lipshitz in the neighborhood of $\hat{x}$ and f is strictly differentiable at $\hat{x}$, then $\bigtriangledown f(\hat{x}) = 0$.*

Given that EPSAs have refining subsequences with probability one, it follows that EPSAs converge to these limit points with probability one. This convergence theory generalizes the result in Theorem 1 in several respects. First, note that if f is continuously differentiable at x then f is Lipshitz in the neighborhood of x and f is strictly differentiable at x [7]. Consequently, Theorems 3 and 4 make weaker assumptions than Theorem 1. Further, our focus on the limit points of EPSAs describes their convergence behavior locally, and hence it is applicable to functions for which continuity properties vary across the search domain.

3.2 Convergence Proofs

In this section we prove the results of the previous section. Recall our assumption that the points sampled by an EPSA lie in a compact set. We begin by showing that there is a subsequence of iterations for which the step lengths go to zero. This proof requires the following two lemmas.

Lemma 1 ([2], Lemma 3.1) *The step length Δ_t is bounded above by a positive constant independent of t.*

Let $\mathcal{X} = \{x \mid \exists i_1, \ldots, i_n \in \{1, \ldots, n\} \text{ s.t. } x = e_1^T x_{i_1}^0 + \ldots + e_n^T x_{i_n}^0\}$, where x_i^0 is the i-th member of the initial population of the EPSA and e_i is the i-th unit vector. The set $\mathcal{X}$ represents the set of points that can be generated by coordinate-wise recombinations of the initial population. Note that we can write $\Delta_t = \Delta_0 \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu}$, where $r_t^i \in \mathbf{Z}^{\geq 0}$ and $\phi_i \in \Phi$. For $i = 1, \ldots, \nu$, let $r_t^{i,\max} = \max_{j=1,\ldots,t} r_j^i$, and note that $\phi_i = \tau_n^i / \tau_d^i$ where $\tau_n^i, \tau_d^i \in \mathbf{Z}^{>0}$. Thus we can define

$$\overline{\tau}_t = (\tau_d^1)^{r_t^{1,\max}} \cdots (\tau_d^\nu)^{r_t^{\nu,\max}}.$$

Now let $\mathcal{S}'$ consist of the set of search directions defined by all $S \in \mathcal{S}$, and consider

$$M_t = \left\{ x + \frac{\Delta_0}{\overline{\tau}_t} \sum_{s \in \mathcal{S}'} Z_s s \mid x \in \mathcal{X}, Z_s \in \mathbf{Z}^{n \times n} \right\}. \tag{4}$$

The set M_t defines a union of meshes, one for every $x \in \mathcal{X}$, that are composed by the lattices spanned by the directions in $\mathcal{S}'$. The following lemma shows that the points in the t-th iteration of an EPSA, X_t, lie in M_t.

Lemma 2 *For all t, $X_t \subset M_t$.*

Proof. Clearly $X_0 \subset \mathcal{X} \subset M_0$. By induction we assume that $X_t \subset M_t$. Consider $x \in X_{t+1}$. This point is either (1) in X_t and hence $x \in M_t \subseteq M_{t+1}$, or (2) it was formed from crossover or mutation or both. In this case we have

$$x = \delta_c[Ax_a^t + (I - A)x_b^t] + \delta_m \Delta_t \overline{s},$$

where (a) $\delta_c, \delta_m \in \{0,1\}$ and $\delta_c + \delta_m \geq 1$, (b) $\overline{s} \in \mathcal{S}'$, (c) $A = \mathrm{diag}(a_1, \ldots, a_n), a_i \in \{0,1\}$, (d) I is the identity matrix, and (e) $x_a^t, x_b^t \in X_t$. It follows that there exists x_a, x_b, Z_s^a and Z_s^b such that

$$\begin{aligned} x &= \delta_c \left[A \left(x_a + \frac{\Delta_0}{\overline{\tau}_t} \sum_{s \in \mathcal{S}'} Z_s^a s \right) + (I - A) \left(x_b + \frac{\Delta_0}{\overline{\tau}_t} \sum_{s \in \mathcal{S}'} Z_s^b s \right) \right] + \delta_m \Delta_0 \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu} \overline{s} \\ &= \delta_c[Ax_a + (I - A)x_b] + \frac{\Delta_0}{\overline{\tau}_t} \sum_{s \in \mathcal{S}'} \delta_c[AZ_s^a + (I - A)Z_s^b]s + \delta_m \frac{\Delta_0}{\overline{\tau}_t} \overline{\tau}_t \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu} \overline{s}. \end{aligned}$$

Note that $\overline{\tau}_{t+1} = \rho_t \overline{\tau}_t$, where $\rho_t \in \{1, \tau_d^1, \ldots, \tau_d^\nu\}$. Let $B_x = \delta_c[Ax_a + (I - A)x_b]$ and $B_s = \delta_c[AZ_s^a + (I - A)Z_s^b]$. Thus we have

$$x = B_x + \frac{\Delta_0}{\overline{\tau}_{t+1}} \left(\sum_{s \in \mathcal{S}'} \rho_t B_s s + \delta_m \rho_t \overline{\tau}_t \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu} I \overline{s} \right).$$

Now $B_x \in \mathcal{X}$, $\rho_t B_s \in \mathbf{Z}^{n \times n}$ and $\overline{\tau}_t \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu} \in \mathbf{Z}$, so $\delta_m \rho_t \overline{\tau}_t \phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu} I \in \mathbf{Z}^{n \times n}$. Thus $x \in M_{t+1}$. ■

The following corollary shows that the intersection of the compact set containing the points generated by the EPSA and M_t is finite.

Corollary 1 *Let Ω' be the compact set that contains the points generated by the EPSA, $\bigcup_{t=1}^{\infty} X_t$. Then $\Omega' \bigcap M_t$ is finite.*

Proof. Recall that for all $s \in \mathcal{S}'$ we have $s \in \mathbf{Q}^n$. Let C be the greatest common divisor of the elements of s for all $s \in \mathcal{S}'$. Thus there exists $z_s \in \mathbf{Z}^n$ such that $s = z_s / C$. We can rewrite M_t as follows:

$$M_t = \left\{ x + \frac{\Delta_0}{\overline{\tau}_t C} \sum_{s \in \mathcal{S}'} Z_s z_s \mid x \in \mathcal{X}, Z_s \in \mathbf{Z}^{n \times n} \right\}.$$

Since Ω' is compact it is also bounded. Thus the projection of Ω' onto the i-th coordinate axis is a closed interval. Now every value in the i-dimension of points in $\Omega' \bigcap M_t$ can be represented as $x_i + \frac{\Delta_0}{\overline{\tau}_t C} w$ for some $w \in \mathbf{Z}$. Since these values are bounded above and below, there are a finite number of distinct values for w that are feasible. Thus the projection of $\Omega' \bigcap M_t$ onto the i-th dimension is finite. It follows that $\Omega' \bigcap M_t$ is finite since this is true for every dimension. ■

Corollary 1 along with Lemmas 1 and 2 are used to prove the following proposition, which ensures that the step length parameter converges to zero if the EPSA either finds improving steps or contracts the step length infinitely often.

Proposition 2 *If the points sampled by an EPSA lie in a compact set and the conditions in steps (20) or (23) are true infinitely often, then* $\liminf_{t\to\infty} \Delta_t = 0$.

Proof. Suppose that $0 < \Delta_{\text{min}} \le \Delta_t$ for all t. The hypothesis that $\Delta_{\text{min}} \le \Delta_t$ for all t means that the sequence $\{\phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu}\}$ is bounded away from zero. We also know from Lemma 1 that the sequence $\{\Delta_t\}$ is bounded above. Thus the sequence $\{\phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu}\}$ is bounded above, from which it follows that the sequence $\{\phi_1^{r_t^1} \cdots \phi_\nu^{r_t^\nu}\}$ is a finite set. Equivalently, the sequences $\{r_t^i\}$ are bounded above and below. Let $r^i_{\text{max}} = \max_{0\le t\le\infty} r_t^i$ and define

$$\overline{\tau}_\infty = (\tau_d^1)^{r^1_{\text{max}}} \cdots (\tau_d^\nu)^{r^\nu_{\text{max}}}.$$

Then we can define a generalized mesh M_∞ using $\overline{\tau}_\infty$ in Equation (4) in place of $\overline{\tau}_t$. Now $\overline{\tau}_t \le \overline{\tau}_\infty$, so $M_t \subseteq M_\infty$. It follows that $X_t \subset M_\infty$ for all t. The analysis from Corollary 1 applies equally well to M_∞, so we know that the intersection of Ω' and M_∞ is finite. Thus there must exist a point $\hat{x}$ for which $x_t = \hat{x}$ for infinitely many t. However, this is a contradiction since we cannot revisit a point in M_∞ infinitely many times. We accept a new mutation step s_t from x_t^* if and only if $f(x_t^*) > f(x_t^* + s_t)$, so there exists N such that for all $t \ge N$, $x_t = \hat{x}$. However, if this is true then we are guaranteed that the condition in step (23) is true infinitely often, which implies that $\Delta_t \to 0$. This gives a contradiction to our assumption that $\Delta_t \ge \Delta_{\text{min}} > 0$. ■

The following lemma ensures that with probability one an EPSA will either find an improving step or it will sample all of the mutation offsets from x_t^*. This result is equivalent to our analysis in Hart [22] which shows that each iteration of a stochastic pattern search algorithm terminates with probability one.

Lemma 3 ([22], Lemma 2) *Let $\mathcal{I}$ be the set of sequences of iterations $\{x_t^*\}$ for an EPSA for which the conditions in steps (20) or (23) are true infinitely often. Then $P(\mathcal{I}) = 1$.*

Proposition 1 follows directly from Lemma 3 and Proposition 2. We now prove Theorems 3 and 4. Our analysis here is similar in spirit to the proofs of Lemma 3.4 and Theorem 3.5 in Audet and Dennis [2].

Proof.[Theorem 3] Let $\hat{S} \in \mathcal{S}$ be a positive spanning set that is used infinitely many times in the refining subsequence; there must be such a set since $\mathcal{S}$ is finite. Let $K' \subseteq K$ denote the subsequence where $\hat{S}$ is used. Since the iterations K' belong to a refining subsequence, it follows that for each $s \in \hat{S}$ and $k \in K'$; either $x_k^* + \Delta_k s$ is infeasible or the step s has been unsuccessfully applied to x_k^*.

Consider a step $s \in \hat{S}$ that can be feasibly applied for infinitely many iterations. From Clarke [7], we have by definition that

$$f^o(\hat{x}; s) = \limsup_{y \to \hat{x}, t \downarrow 0} \frac{f(y + ts) - f(y)}{t} \geq \limsup_{k \in K'} \frac{f(x_k^* + \Delta_k s) - f(x_k^*)}{\Delta_k}.$$

Note that f is Lipshitz near $\hat{x}$, so f must be finite near $\hat{x}$. Since s can be feasibly applied infinitely many times, infinitely many terms of the right quotient sequence are defined. All of these terms are nonnegative because the steps are unsuccessful, so it follows that $f^o(x; s) \geq 0$. ■

Proof.[Theorem 4] From Theorem 3 we know that there exists $S \in \mathcal{S}$ such that $f^o(\hat{x}; s) \geq 0$ for all $s \in S$. For all $w \in \mathbf{R}^n$ we can write w as a nonnegative linear combination of elements of S, from which it follows that $\bigtriangledown f(\hat{x})^T w \geq 0$. The same construction for $-w$ shows that we have $- \bigtriangledown f(\hat{x})^T w \geq 0$, so $\bigtriangledown f(\hat{x})^T w = 0$. ■

4 Constrained Analysis

This section describes how the previous analysis of EPSAs can be extended to solve problem (3) for general linearly constrained problems. The central difference of this analysis are restrictions that guarantee that the search directions reflect the geometry of the constraint boundary when the EPSA converges to points near the boundary. These restrictions are needed to ensure that good search directions are available for the mutation operator near the boundary.

The approach that we take here is similar to the approach described by Michalewicz and Attia [26, 27] to the extent that EPSAs adapt their mutation operator based on the properties of the constraint boundary. However, EPSAs may generate infeasible points that are simply rejected; infeasible points are never evaluated. This method is similar to rejection methods commonly used with EP and ESs for bound constraints. One important difference is that because EPSAs use a finite number of mutation offsets, they are guaranteed to shrink the step-length parameter after generating a finite number of infeasible mutation steps. Also note that since EPSAs are restricted to coordinate-wise crossover, the crossover operator is very likely to generate infeasible points for a general linearly constrained domain, so its utility for this class of problems is questionable.

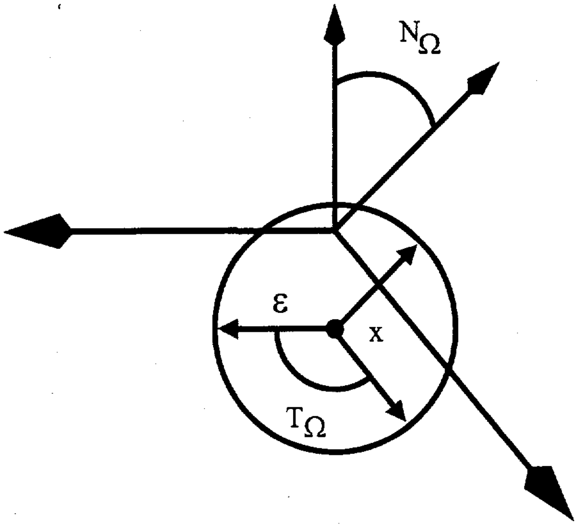

Figure 2: Illustration of the cones $T_\Omega(x,\epsilon)$ and $N_\Omega(x,\epsilon)$ for the constraints are $y \leq 0$, $x+y \leq 0$, the point $x=(0,-1)$ and $\epsilon = 1.25$.

The geometry of the constraint boundary can be described using the following definitions of generalized tangent and normal cones. For all $\epsilon > 0$ and $x \in \Omega$, let $V(x,\epsilon)$ be the normals to the faces of the boundary that are within distance ϵ of x. The generalized normal cone, $N_\Omega(x,\epsilon)$, is the cone generated by the vectors in $V(x,\epsilon)$. That is,

$$N_\Omega(x,\epsilon) = \left\{ v \mid v = \sum_{i=1}^{|V(x,\epsilon)|} \lambda_i v_i, \lambda_i \geq 0, v_i \in V(x,\epsilon) \right\}.$$

The generalized tangent cone is $T_\Omega(x,\epsilon) = \{v \in \mathbf{R}^n \mid \forall w \in N_\Omega(x,\epsilon), v^T w \leq 0\}$. Figure 2 illustrates the cones near a two-dimensional boundary. To accomodate the geometry of the constraint boundary, we use a mapping $S_\epsilon : \Omega \to \mathcal{S}$, for $\epsilon > 0$, to select a matrix that defines mutation offsets for a given point. Given $\epsilon > 0$, we say that S_ϵ is an *ϵ-conforming mapping* if for all $x \in \Omega$ some subset of columns of $S_\epsilon(x)$ generates $T_\Omega(x,\epsilon)$.

For example, consider the point x in Figure 2. The two vectors that generate $T_\Omega(x,\epsilon)$ are $(-1,0)^T$ and $(1,-1)^T$. An ϵ-conforming mapping is free to select any matrix $S \in \mathcal{S}$ that contains these columns. The additional column in S is simply required to form a positive basis for $\mathbf{R}^n$. For example, the vector $(1,1)$ suffices.

By using an ϵ-conforming mapping, we guarantee that the set of mutation directions used by an EPSA contains directions that generate points in a feasible direction. Thus if a point is not a constrained stationary point, then we can reduce the step length to find a descent direction. Further, when the point is not near the boundary, then the set of mutation directions are simply a positive basis, so our convergence result for the unconstrained case applies. The following proposition is an immediate corollary of Theorem 3.

Proposition 3 *Given $\epsilon > 0$, let $\hat{x}$ be the limit of a refining subsequence for an EPSA that selects $\overline{S}$ with an ϵ-conforming mapping, S_ϵ, at x_t^*. If f is Lipschitz near $\hat{x}$ then $f^o(\hat{x}; s) \geq 0$ for all directions $s \in S_\epsilon(\hat{x})$ that generate $T_\Omega(\hat{x}, \epsilon)$.*

Note that this proposition simply requires that the mutation offsets about the best point in the population be adapted to ensure convergence. This has the advantage of minimizing the changes made to the generic EA framework to enable convergence for linearly constrained problems. However, in Section 6 we discuss whether this method of formulation EPSAs is well-suited for a population-based search method like EAs.

Recall that a constrained stationary point $\hat{x}$ for problem (3) is a KKT point [18], so for linear inequalities it sufficies to have $\bigtriangledown f(\hat{x})^T w \geq 0$ for all $w \in T_\Omega(\hat{x}, 0)$, and $- \bigtriangledown f(\hat{x}) \in N_\Omega(\hat{x}, 0)$. The following proposition describes our main convergence result for EPSAs on problem (3).

Theorem 5 *Given $\epsilon > 0$, let $\hat{x}$ be the limit of a refining subsequence for an EPSA that selects $\overline{S}$, with an ϵ-conforming mapping, S_ϵ, at x_t^*. There exists $\epsilon^* > 0$ such that if $\epsilon^* > \epsilon$ and f is strictly differentiable at $\hat{x}$, then $\hat{x}$ is a KKT point.*

Proof. If $\hat{x}$ is in the interior of Ω, then the ϵ-conforming mapping selects positive spanning sets for any $\epsilon > 0$. Thus our result follows from Theorem 4 and the definition of a KKT point.

Otherwise, note that since there are a finite number of faces of Ω, there exists $\epsilon^* > 0$ such that $T_\Omega(\hat{x}, \epsilon) = T_\Omega(x, 0)$ for all $\epsilon^* > \epsilon > 0$ and all x on the boundary of Ω. From Proposition 3 we know that $\bigtriangledown f(\hat{x})^T s \geq 0$ for all $s \in S_\epsilon(\hat{x})$ that generate $T_\Omega(\hat{x}, \epsilon)$. Now every $w \in T_\Omega(\hat{x}, 0) = T_\Omega(x, \epsilon)$ is a nonnegative linear combination of these vectors, so $\bigtriangledown f(\hat{x})^T w \geq 0$. But $(- \bigtriangledown f(\hat{x})^T) w \leq 0$, so $- \bigtriangledown f(\hat{x}) \in N_\Omega(\hat{x}, 0)$. ■

4.1 Construction of Patterns

The general framework used by this convergence theory begs the question of whether it is possible to practically define an ϵ-conforming mapping that can be used by EPSAs. Fortunately, Lewis and Torczon [24] describe a construction for pattern search methods that can be used to determine a positive basis that generates $T_{\Omega}(x, \epsilon)$. The description of this construction is beyond the scope of our presentation here, though we note that this method of constructing a positive basis makes the assumption that for all x on the boundary of Ω, $V(x, \epsilon)$ is comprised of a linearly independent set of vectors. Thus at any point on the boundary the entire set of active constraints must be tight, which implies that the set of linear constraints is nondegenerate. Further, note that the construction described by Lewis and Torczon [24] provides a method for estimating ϵ^* throughout the course of optimization. The value of ϵ^* depends upon the linear constraints, so it is important to determine this value for each problem.

4.2 The Bound-Constrained Case

We now consider the special case where the linear constraints simply define bound constraints (i.e. A is a diagonal matrix). Theorem 5 clearly generalizes our previous result in Theorem 2. In the case of bound constraints, we know *a priori* the possible generators of $T_{\Omega}(x, \epsilon)$ and $N_{\Omega}(x, \epsilon)$. For any $x \in \Omega$ and $\epsilon > 0$, the cone $N_{\Omega}(x, \epsilon)$ is generated by some subset of the coordinate vectors $\pm e_i$. Thus we can simply use the positive basis that contains all coordinate vectors. This choice includes generators for all possible $T_{\Omega}(x, \epsilon)$, so it is ϵ-conforming for all $\epsilon > 0$ and for all t.

Further, note that Theorem 5 is applicable when some of the search dimensions are unconstrained: $l_i, u_i = \pm\infty, i \in \{1, \ldots, n\}$. As Lewis and Torczon [24] note, we can make a more parsimonious choice of a spanning set in this case. Let $x_{i_1}, \ldots, x_{i_r}$ be the variables with either a lower or upper bound. Then the positive spanning set includes $\pm e_{i_1}, \ldots, \pm e_{i_r}$ as well as a positive basis for $\{v \mid v^T z = 0, z = \sum_{j=1}^{r} \lambda_j e_{i_j}, \lambda_j \in \mathbf{R}\}$; the positive basis for this set can have as few as $n - r + 1$ elements. This spanning set is also independent of t, and it requires at most $n + r + 1$ function evaluations to perform a contraction of the step length parameter.

4.3 Generation of Initial Points

Our analysis of EPSAs for general linearly constrained problems assumes that the initial population consists of feasible points in the polyhedron defined by inequality constraints. Following the common design of EAs, it is desirable to generate points uniformly at random within Ω. Unfortunately, it is likely that an efficient process for generating this distribution is not possible since the problem of computing the exact volume of a polytope (a bounded polyhedron) is #P-complete [12]. For example, one obvious method for generating points in Ω with a uniform distribution is to uniformly generate points in a box (for a polytope) or half-plane (for a polyhedron) and reject points that do not lie in Ω. However, it is easy to construct problems for which the fraction of the region that lies in Ω is arbitrarily small, so the efficiency of this approach is poor.

Methods for generating points uniformly on polytopes are discussed by Devroye [11], Leydold and Hörmann [25], and Rubin [28, 29]. The complexity of these methods is at least polynomial in the dimension and the number of vertices of the polytope defined by the linear inequalities. Since the number of points that are needed to setup an EPSA is typically quite small, the methods described by Rubin [29] and Leydold and Hörmann [25] seem most appropriate. Rubin describes a random walk on a polytope that asymptotically samples the feasible domain uniformly. Leydold and Hörmann describe a sweep-plane method that recursively sweeps a plane through the polytope until the plane has swept through a fixed percentage of the volume, which is chosen at random.

Although these approaches avoid an expensive setup phase, they still may take a long time for high dimensional problems or for polytopes with many vertices. Consequently, we believe that it is worth considering methods that can quickly generate points from a nonuniform distribution that is not too biased. For example, we have developed a randomized version of the phase one algorithm that is commonly used to find a feasible point in linear programming algorithms (e.g. see Gill, Murray and Wright [18], Section 5.7). Our randomized approach generates a point in an enclosing box (half-plane) for the polytope (polyhedron). Then the randomized phase one algorithm iteratively choses a violated constraint and moves the point to make it feasible (while maintaining the feasibility of all feasible constraints). This process is very fast, though its empirical utility remains to be demonstrated.

5 Deterministic Analysis

Our analysis in the previous sections ensures convergence to interesting limit points with probability one. However, a stochastic convergence guarantee is not necessary simply because of the fundamentally stochastic nature of the EPSAs defined in Figure 1. Although this definition of EPSAs comes closest to capturing the basic algorithmic framework of generic EAs, with only modest restrictions we can ensure deterministic convergence of $\{x_t^*\}$ to the same class of limit points. Consider the following assumption on the properties of an EPSA.

Assumption 1 *Consider an EPSA for which*

1. *The* **selection** *function always includes the point x_t^* in $\overline{X}$*

2. *The* **crossover** *function is applied to generate $N_c < N$ new individuals*

3. *A mutation step $s_i \in \overline{S}$ is applied to the point x_t^* (without crossover), where the steps in $\overline{S}$ are sampled in a fixed order (which may be randomly chosen).*

This assumption ensures that in every iteration a different mutation step from the point x_t^* is generated. Note that the updates to Δ_t fundamentally depend upon the success or failure of these mutation steps. Thus this assumption ensures that every iteration of an EPSA makes progress in the optimization, both by avoiding mutation steps that have already been evaluated and by ensuring that these mutation steps occur all the time. The following proposition follows immediately from this assumption.

Proposition 4 *If the sequence $\{x_t^*\}$ is generated by an EPSA that satisfies Assumption 1, then there exists a refining subsequence of $\{x_t^*\}$.*

Proof. If Assumption 1 is satisfied, then in every iteration a unique mutation offset of x_t^* is considered. Thus the conditions in steps (20) or (23) are true infinitely often. It follows from Proposition 2 that $\liminf_{t\to\infty} \Delta_t = 0$, so there exists a refining subsequence of $\{x_t^*\}$. ■

We argue that Assumption 1 imposes relatively weak restrictions on an EPSA. It is easy to satisfy Assumption 1.1 using any common selection mechanism that is applied using stochastic-universal-selection [5]. Further, crossover operators are commonly applied in EAs with probability less than one, so Assumption 1.2 is not particularly restrictive.

Satisfying Assumption 1.3 requires a modest algorithmic change. An index array for $\overline{S}$ needs to be constructed and randomly shuffled, which requires $O(n)$ work. Given this shuffled array, the mutation steps for x_t^* are iteratively selected by simply taking the next value in the array. Furthermore, if by chance two or more points in the population equal x_t^*, they can evaluate distinct mutation steps and accelerate the convergence of the EPSA.

6 Discussion

This paper adapts Audet and Dennis' [2] recent observations on the requirements for convergence in pattern search algorithms to significantly extend and simplify our previous convergence analysis of EPSAs. We have weakened the continuity assumptions made on the objective function. and we have presented a convergence theory that allows for different continuity properties in different neighborhoods of the search domain. Our convergence analysis encompasses our previous analysis for unconstrained and bound constrained problems, and it provides a general framework for extending it to general linearly constrained problems.

This analysis is significant because it exactly characterizes the convergence properties of an adaptive EA on a broad class of nonconvex objective functions. Consequently, this convergence theory provides a rigorous justification for the use of adaptive EAs in a wide range of problems. Although other adaptive methods may be used for other classes of EAs, we expect that this analysis will help illustrate the type of adaptation that is needed to ensure robust convergence. For example, the adaptation of the mutation offsets near a constraint boundary reflects fundamental issues that will need to be tackled by any EA.

The principle challenge that remains unresolved in this work is the design of effective EPSAs for linearly constrained problems. We have noted that Lewis and Torczon [24] provide an algorithm that can be used to select mutation offsets for EPSAs on linearly constrained problems. However, this framework needs to be evaluated and refined for use within the context of EPSAs.

A fundamental weakness of our formulation of linearly constrained EPSAs is that the mutation offsets are only tailored to the local geometry about the point x_t^*. This ignores the fact that the geometry is likely to be very different near other points in the population, thereby limiting the overall effectiveness of a population-based method like EPSAs. We believe that this framework for defining EPSAs can be extended to define EPSAs that locally adapt the geometry of mutation offsets while retaining the general convergence theory. Unfortunately, this will certainly complicate the design of EPSAs, removing them further from the design of canonical EAs.

Two other algorithmic details will probably also need to be addressed to make EPSAs practical for linearly constrained problems. First, it is often quite desirable to move to a constraint boundary when the search approaches the boundary of the feasible domain (e.g. see [26]). Unfortunately, the search method employed by EPSAs does not allow this type of a move, and the set of feasible points generated by an EPSA will not generally contain points on the boundary of Ω.

The second algorithmic issue is the design of effective recombination operators. As we noted earlier, the utility of coordinate-wise crossover operators is questionable for linearly constrained problems. It is easy to construct a problem for which a crossover operator will only generate feasible points if the two initial points are very close in the search domain. However, recombination operators are generally viewed as methods for generating "global" steps that combine subsolutions from different parts of the search domain. To acheive this functionality, new crossover operators will need to be developed for EPSAs that mimic blending crossover operators (e.g. see [14, 31]).

Finally, we note that experiments will be needed to evaluate the impact of the restrictions imposed by Assumption 1. An argument for *not* satisfying these assumptions is that EAs are really best suited for global optimization, and hence it is appropriate to employ mechanisms that only weakly encourage convergence to a local optima. For example, we have previously demonstrated the empirical utility of EPSAs on standard global optimization test problems and on a real-world application [20, 21, 23]. Although EPSAs performed about as well as other EAs on these applications (and sometimes much better), encouraging convergence to local optima might limit their ability to perform a global search. It should be noted, however, that the rate of convergence of EPSAs is likely to be very poor for large-scale problems, so the mechanisms required by Assumption 1 will likely be necessary in this case.

Acknowledgements

We thank John Dennis and Tammy Kolda for their helpful discussions. This work was performed at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

References

[1] Audet, C. (1998). Convergence results for pattern search algorithms are tight, *Tech. Rep. TR98-24*, Rice University, Department of Computational and Applied Mathematics.

[2] Audet, C., and J. E. Dennis Jr. (2000). Analysis of generalized pattern searches, *Tech. Rep. TR00-07*, Rice University, Department of Computational and Applied Mathematics.

[3] Bäck, T., and H.-P. Schwefel (1993). An overview of evolutionary algorithms for parameter optimization, *Evolutionary Computation*, *1*(1), 1–23.

[4] Bäck, T., F. Hoffmeister, and H.-P. Schwefel (1991). A survey of evolution strategies, in *Proc. of the Fourth Intl. Conf. on Genetic Algorithms*, edited by R. K. Belew and L. B. Booker, pp. 2–9, Morgan-Kaufmann, San Mateo, CA.

[5] Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm, in *Proc. of the Second Intl. Conf. on Genetic Algorithms*, pp. 14–21.

[6] Calamai, P. H., and Moré (1987). Projected gradient methods for linearly constrained problems, *Mathematical Programming*, *39*, 93–116.

[7] Clarke, F. H. (1990). *Optimization and Nonsmooth Analysis*, vol. 5, SIAM Classics in Applied Mathematics, Philadelphia, PA.

[8] Conn, A. R., N. I. M. Gould, and P. L. Toint (1988). Global convergence of a class of trust region algorithms with simple bounds, *SIAM J Numerical Analysis*, *25*, 433–460.

[9] Davis, C. (1954). Theory of positive linear dependence, *American Journal of Mathematics*, pp. 733–746.

[10] Dennis, J. J., and R. B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall.

[11] Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

[12] Dyer, M. E., and A. M. Frieze (1988). On the complexity of computing the volume of a polyhedron, *SIAM J Comput*, *17*(5), 967–974.

[13] Eldred, M. S. (1998). Optimization strategies for complex engineering applications, *Tech. Rep. SAND98-0340*, Sandia National Laboratories.

[14] Eshelman, L. J., and J. D. Schaffer (1993). Real-coded genetic algorithms and interval schemata, in *Foundations of Genetic Algorithms 2*, edited by L. D. Whitley, pp. 187–202, Morgan-Kauffmann, San Mateo, CA.

[15] Fogel, D. B. (1994). An introduction to simulated evolutionary optimization, *IEEE Trans Neural Networks*, *5*(1), 3–14.

[16] Fogel, D. B. (1995). *Evolutionary Computation*, IEEE Press, Piscataway, NJ.

[17] Folland, G. B. (1984). *Real Analysis - Modern Techniques and Their Applications*, John Wiley and Sons.

[18] Gill, P. E., W. Murray, and M. H. Wright (1981). *Practical Optimization*, Academic Press.

[19] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Co., Inc.

[20] Hart, W. E. (1998). On the application of evolutionary pattern search algorithms, in *Proc Evolutionary Programming VII*, pp. 303–312, Springer-Berlin, New York.

[21] Hart, W. E. (1999). Comparing evolutionary programs and evolutionary pattern search algorithms: A drug docking application, in *Proc. Genetic and Evolutionary Computation Conf*, pp. 855–862.

[22] Hart, W. E. (2000). A convergence analysis of unconstrained and bound constrained evolutionary pattern search, *Evolutionary Computation*, (to appear).

[23] Hart, W. E., and K. Hunter (1999). A performance analysis of evolutionary pattern search with generalized mutation steps, in *Proc Conf Evolutionary Computation*, pp. 672–679.

[24] Lewis, M., and V. Torczon (1998). Pattern search methods for linearly constrained minimization, *SIAM J Opt*, (to appear).

[25] Leydold, J., and W. Hörmann (1998). A sweep-plane algorithm for generating random tuples in simple polytopes, *Mathematics of Computation*, *67*, 1617–1635.

[26] Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed., Springer-Verlag.

[27] Michalewicz, Z., and N. Attia (1994). Evolutionary optimization of constrained problems, in *Proc. of the Third Annual Conf. on Evolutionary Programming*, pp. 98–108.

[28] Rubin, P. A. (1984). Generating random points in a polytope, *Communications in Statistics - Simulation*, *13*(3), 375–396.

[29] Rubin, P. A. (1985). Short-run characteristics of samples drawn by random walks, *Communications in Statistics - Simulation*, *14*(2), 473–490.

[30] Rudolph, G. (1997). *Convergence Properties of Evolutionary Algorithms*, Kovač.

[31] Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*, John Wiley & Sons, New York.

[32] Torczon, V. (1991). On the convergence of the multidirectional search algorithm, *SIAM J. Optimization*, *1*, 123–145.

[33] Torczon, V. (1997), On the convergence of pattern search methods, *SIAM J Optimization*, 7(1), 1–25.