On Gradient Based Local Search Methods in Unconstrained Evolutionary Multi-objective Optimization

Pradyumn Kumar Shukla

Institute for Numerical Mathematics
Dresden University of Technology
Dresden, D-01062, Germany
pradyumn.shukla@mailbox.tu-dresden.de

Abstract. Evolutionary algorithms have been adequately applied in solving single and multi-objective optimization problems. In the single-objective case various studies have shown the usefulness of combining gradient based classical methods with evolutionary algorithms. However there seems to be limited number of such studies for the multi-objective case. In this paper, we take two classical methods for unconstrained multi-optimization problems and discuss their use as a local search operator in a state-of-the-art multi-objective evolutionary algorithm. These operators require gradient information which is obtained using finite difference method and using a stochastic perturbation technique requiring only two function evaluations. Computational studies on a number of test problems of varying complexity demonstrate the efficiency of resulting hybrid algorithms in solving a large class of complex multi-objective optimization problems. We also discuss a new convergence metric which is useful as a stopping criteria for problems having an unknown Pareto-optimal front.

1 Introduction

Multi-objective optimization is a rapidly growing area of research and application in modern-day optimization. There exist a plethora of non-classical methods which follow some natural or physical principles for solving multi-objective optimization problems, see for example the book by [5]. On the other hand a large amount of studies have been devoted to develop classical methods for solving multi-objective optimization problems ([8]).

Evolutionary algorithms use stochastic transition rules using crossover and mutation search operators to move from one solution to another. In this way global structure of search space is exploited. Classical methods, on the other hand, usually use deterministic (usually gradient based) transition rules to move from one solution to another. Classical methods effectively use local information thus ensuring fast convergence. This however comes up at the cost of requiring gradient or Hessian information which requires a large number of function evaluations. Hence one sees that there is a trade-off between fast convergence

S. Obayashi et al. (Eds.): EMO 2007, LNCS 4403, pp. 96–110, 2007.

and number of function evaluations. Hybrid implementations thus continue to be developed and tested (see for example [1,2,10,11]).

In this contribution we take two classical gradient based Pareto front generating methods and use their search principles as mutation operators in a state-of-the-art multi-objective evolutionary algorithm to create a powerful hybrid multi-objective metaheuristics algorithm. We demonstrate their efficiency in solving real valued differentiable problems of varying complexity.

This paper is structured as follows. The next section present an overview of various classical generating methods and the gradient estimation technique, while the third section presents the simulation results. Conclusions as well as extensions which emanated from this study are presented in the end of this contribution.

2 Classical Generating Methods

For the present study we take two classical algorithms and use their search operators as mutation operator in the elitist non-dominated sorting GA or NSGA-II developed by [6]. The gradients of objective functions are numerically computed by two methods: one-sided finite difference method and a stochastic perturbation method. These gradient estimation methods and the classical search operators used in this study are described in this section.

2.1 Gradient Estimation Methods

In almost all classical algorithms (for both single and multi-objective problems) the gradient of a function (say in general h) are required. The standard approach for estimating the gradient is the Finite Difference (FD) method (one-sided or two-sided). Let $\mathbf{e}_i$ denote a unit vector in the i^{th} direction, then for a variable (say $\mathbf{x}$) of dimension n the one-sided FD method of gradient estimation requires $(n+1)$ function evaluations and is given by

$$g_i(\mathbf{x}) = \frac{f(\mathbf{x} + c\mathbf{e}_i) - f(\mathbf{x})}{c},$$

This is costly in terms of function evaluations (of the order $O(n)$). The Simultaneous Perturbation (SP) method [16] on the other hand requires function evaluation independent of n. The one-sided SP method required only *one* additional function measurement and is thus $O(1)$) as follows

$$g_i(\mathbf{x}) = \frac{f(x + c\Delta) - f(x)}{c\Delta_i},$$

where the i^{th} component of the gradient is denoted by $g_i(\mathbf{x})$, Δ is a n dimensional vector (Δ_i is its i^{th} component) of random perturbations satisfying certain statistical conditions ([16]). A simple (and theoretically valid) choice for each component of Δ is to use a Bernoulli distribution ± 1 with probability of 0.5 for each ± 1 outcome. The step size c at each iteration (denoted by c_k) is given as $c_k = c_0/(k+1)^\gamma$. Practically effective (and theoretically valid [16]) values of c_0, γ are 0.001 and 1/6 which are used here.