European Journal of Operational Research 235 (2014) 38–46

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Continuous Optimization

Line search methods with guaranteed asymptotical convergence to an improving local optimum of multimodal functions

Douglas Alexandre Gomes Vieira *, Adriano Chaves Lisboa

ENACOM – Handcrafted Technologies, Belo Horizonte Technology Park, Rua Prof. José Vieira de Mendonça, 770, Belo Horizonte/MG, Brazil

ARTICLE INFO

Article history:
Received 23 February 2013
Accepted 25 December 2013
Available online 6 January 2014

Keywords:
Nonlinear programming
Line search
Golden section method
Brent's algorithm
Bisection
Multimodal functions

ABSTRACT

This paper considers line search optimization methods using a mathematical framework based on the simple concept of a v-pattern and its properties. This framework provides theoretical guarantees on preserving, in the localizing interval, a local optimum no worse than the starting point. Notably, the framework can be applied to arbitrary unidimensional functions, including multimodal and infinitely valued ones. Enhanced versions of the golden section, bisection and Brent's methods are proposed and analyzed within this framework: they inherit the improving local optimality guarantee. Under mild assumptions the enhanced algorithms are proved to converge to a point in the solution set in a finite number of steps or that all their accumulation points belong to the solution set.

1. Introduction

This work investigates the solution of the line search optimization problem

$$\begin{aligned} &\text{minimize} \quad f(\alpha) = F(a + \alpha d) \\ &\text{subject to} \quad 0 \leqslant \alpha \leqslant \alpha_{\max} \end{aligned} \tag{1}$$

where $\alpha \in \mathbb{R}, \alpha_{\max}$ is the upper bound for $\alpha, F : \mathbb{R}^n \mapsto \mathbb{F} \subseteq \mathbb{R}$ is an arbitrary function (including discontinuous function), $f : \mathbb{R} \mapsto \mathbb{R}$ is defined for a given point $a \in \mathbb{R}^n$ and a non-null (usually descent) direction $d \in \mathbb{R}^n$.

A solution of the line search problem (1) is typically used inside a search direction optimization algorithm to solve

$$\begin{aligned} &\text{minimize } F(x) \\ &\text{subject to } x_{\min} \leqslant x \leqslant x_{\max} \end{aligned} \tag{2}$$

where $x \in \mathbb{R}^n, x_{\min}$ and $x_{\max}$ are the respective lower and upper bounds for x. In this application, a is the point where the problem's oracle is queried and d is the search direction. This leads to the iterative update $x_{k+1} = x_k + \alpha_k^\star d_k$ where

$$\alpha_k^\star = \arg\min_\alpha F(x_k + \alpha d_k) : 0 \leqslant \alpha \leqslant \alpha_{\max} \tag{3}$$

$$\alpha_{\max} = \max \alpha : x_{\min} \leqslant x_k + \alpha d_k \leqslant x_{\max} \tag{4}$$

and d_k is defined according to the chosen method.

In fact, there is a vast range of exact and inexact methods to solve (1). The simplest approximate solution to the line search problem used inside a search direction algorithm is a constant step size $\alpha_k^\star$. If this constant step is set too small, a slow convergence rate will take place. If it is too large, the search direction algorithm may diverge. This procedure does not guarantee convergence, even when F is assumed to be strictly convex. To overcome these drawbacks, reducing stepsize rules may be employed (Bertsekas, 2008), which guarantee global optimality and termination if F is convex.

Another popular technique is the Armijo's rule, also known as successive stepsize reduction (Armijo, 1966; Shi & Shen, 2005). The Goldstein's test (Goldstein, 1965) adds the condition that the step size is not too small if the Goldstein's rule is verified. If the cost to evaluate the derivative of F is small, the Wolf's test (Wolfe, 1969) can be considered. Some algorithms derived from these tests can be found in the work of Yuan (2010). Non-monotone variations of these techniques are presented in Hu, Huang, and Lu (2010) and Yu and Pu (2008).

Based on interval reductions, the bisection line search (Bertsekas, 2008; Luenberger & Ye, 2010; Bazaraa & Shetty, 2006) removes half of the search interval at each iteration. However, it relies on the gradient of f and, hence, on the differentiability of f. Based on two points information, the golden section algorithm reduces the confidence interval to achieve convergence (Kiefer, 1953; Avriel & Wilde, 1966). Section techniques based on curve fitting have also been addressed, like the Brent's algorithm (Brent, 1973).

This work is mainly interested in section like algorithms. The concept of a v-pattern is explored in this paper to derive some novel line search strategies. Moreover, it is proven that the golden

* Corresponding author. Tel.: +55 31 3401 1043.
E-mail addresses: douglas.vieira@enacom.com.br (D.A.G. Vieira), adriano.lisboa@enacom.com.br (A.C. Lisboa).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.12.041

section algorithm converges to a local minimum for any line search problem (1), without the need of function unimodality. However, as it is well known by practitioners, the algorithm output can be worse than the starting point.

Based on these considerations, this work presents a framework to derive line search section methods for arbitrary functions f, with the guarantee of always keeping a local minimum of (1) whose function value is no greater than $f(0)$. These ideas are applied to the Golden Section, bisection and Brent's algorithms. The enhanced algorithms are used as a block in a general algorithm, and the converge of this algorithm is proved under mild conditions.

In order to contextualize the improvements proposed in this paper, a brief overview of line search methods is given in the next section.

2. Line search methods overview and state of the art

There is no line search method that is the best for all classes of problems, since each one was developed to best explore some conditions. Therefore, the art of matching problems and solvers is a fundamental step in practical optimization. This section describes the most notable line search methods, emphasizing advantages and shortcomings of each one.

2.1. Constant step

A constant step requires no oracle queries: just choose a constant $\alpha_k^\star, \forall k$, where the line search problem is actually bypassed. However, convergence guarantees of a search direction algorithm with constant step are typically problem dependent. Since a search direction algorithm performs the iterative update $x_{k+1} = x_k + \alpha_k^\star d_k$, a constant step $\alpha_k^\star$ implies asymptotical convergence whenever $d_k \to 0$ as $k \to \infty$, considering also that $d_k = 0$ is an optimality condition.

Moreover, it is possible to build a continuously differentiable strictly convex function (see Fig. 1)

$$F(x) = \begin{cases} \frac{3(1+x)^2}{4} - 2x - 1 & ,x < -1 \\ x^2 & ,|x| \leqslant 1 \\ \frac{3(1-x)^2}{4} + 2x - 1 & ,x > 1 \end{cases} \quad (5)$$

such that, for $d_k = -\nabla F(x_k)$ and $\alpha_k^\star = 1$, the objective function decreases asymptotically, i.e. $F(x_{k+1}) < F(x_k), \forall k$, to a non-optimal value by getting trapped at $x_{k+1} = -x_k$ (Bazaraa & Shetty, 2006). The respective sequence is not convergent for $x_0 \neq 0$. This result can be intriguingly stated: taking a sequence of constant non null steps decreasing the objective function, towards shrinking-length directions that are null only at optimal points, does not imply convergence to a finite optimal point, even for continuously differentiable strictly convex problems.

2.2. Backtracking

A typical theoretical guarantee of search direction methods is $F(x_k + \epsilon d_k) < F(x_k)$ or, equivalently, $f(\epsilon) < f(0)$ for a sufficient small $\epsilon > 0$. However, a small step implies slow convergence rate of the search direction algorithm. Conversely, a large step can make the algorithm diverge or converge to a non-local minimum, as previously seen. The backtracking strategy fundamentally reduces a large step size until some conditions which guarantee convergence are met. Just like a constant step size, this strategy does not actually solve the line search problem (1).

2.2.1. Stop conditions

Armijo's condition (Armijo, 1966) for a point $\alpha^\star > 0$ can be written as

$$\begin{aligned} & f(\alpha^\star) \leqslant \tilde{f}(\alpha^\star) \\ & f(\eta\alpha^\star) > \tilde{f}(\eta\alpha^\star), \; \eta > 1 \end{aligned} \quad (6)$$

where $\tilde{f}$ is a variation of the linear approximation at $\alpha = 0$ given by

$$\tilde{f}(\alpha) = f(0) + \epsilon \nabla f(0) \alpha \quad (7)$$

for a fixed $\epsilon \in (0, 1)$. The upper bound condition considers $\nabla f(0) < 0$, otherwise it could not be satisfied by a convex function f. Hence, it guarantees $f(\alpha^\star) < f(0)$, which must be satisfied at least at a sufficient small $\alpha^\star$. Goldstein proposed a stricter condition (Goldstein, 1965) where $\epsilon \in (0, 1/2)$ and $\eta = (1-\epsilon)/\epsilon$. Considering backtracking algorithms, the lower bound condition is useless to be tested. However, $1/\eta$ can be easily related to shrinking rate of the step length in backtracking using α_{max} as starting point, so that Armijo's lower bound condition follows naturally.

2.3. Bisection

Consider a line search problem (1) where f is continuously differentiable and let $\alpha_2 = (\alpha_1 + \alpha_3)/2$ where $[\alpha_1, \alpha_3] \subseteq [0, \alpha_{max}]$. Let $\Delta = \alpha_3 - \alpha_1$ be the interval length. A strictly positive derivative $\nabla f(\alpha_2)$ implies $\exists \epsilon \in (0, \Delta/2)$ such that $f(\alpha_2 - \epsilon) < f(\alpha_2)$, at least for an infinitesimal $\epsilon > 0$, so that $(\alpha_2, \alpha_3]$ can be cut out. Conversely, a strictly negative derivative allows cutting out the subinterval $[\alpha_1, \alpha_2)$. This process can be carried out until the localizing interval becomes arbitrarily small. A null derivative $\nabla f(\alpha_2) = 0$ is a necessary optimality condition for α_2 and, hence, a natural stop criterion. Therefore, the convergence is guaranteed by construction: half of the localizing interval is cut out in each iteration.

Furthermore, the convergence rate is problem independent: $\Delta_k/\Delta_{k-1} = 2^{-1}$ where Δ_k is the localizing interval length at iteration k, so that $\Delta_k/\Delta_0 = 2^{-k}$. These properties imbue some robustness to the bisection method, nevertheless, $\nabla f(\alpha_2) = 0$ is a necessary but not a sufficient optimality condition.

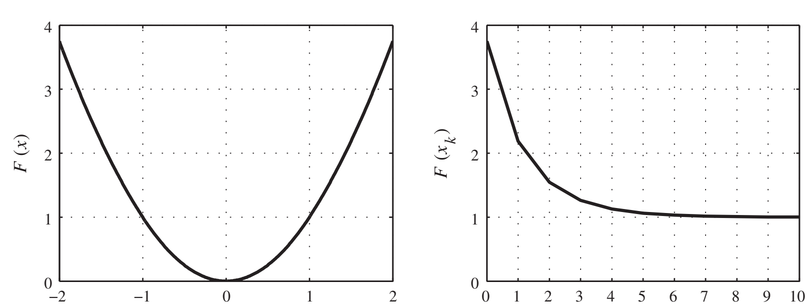

Fig. 1. Instance of (5) (left) where a constant step $\alpha^\star = 1$ line search asymptotically stalls at $F(x) = 1$ (right), for any starting point $x_0 \leqslant -1$ or $x_0 \geqslant 1$ (e.g. $x_0 = 2$ in this case). Note that $F(x_{k+1}) < F(x_k), \forall k$.