A Finite Step-Size Procedure for the F-W Method

LAZAROS P. MAVRIDES

Morgan Guaranty Trust Co., New York, New York 10015

Received December 11, 1978

The F-W Method for the NLP problem with linear constraints requires in its one-dimensional subproblem an exact solution, which cannot be obtained by a finite search procedure. Goldstein has used in unconstrained optimization a one-dimensional subproblem for which any point within certain real-line intervals is an acceptable solution. A similar subproblem can be applied to the F-W Method. A search procedure for such a subproblem is given and proven finite. A stopping rule for the algorithm is also given and proven $\epsilon$-convergent to a K-T point, thus making the algorithm computationally implementable. Numerous applications to a complex banking problem have shown excellent results.

1. INTRODUCTION

We consider the problem

$$\text{maximize } f(x),$$

$$\text{subject to } Ax \geq b,$$

where the objective function $f$ is continuously differentiable, the constraint matrix $A$ is $M \times N$, and the feasible region $F^o$ is compact. The F-W Method (Frank and Wolfe [2]) at each iteration seeks first a direction $d = z^* - x$, where $x$ is a point of iteration and $z^*$ is an optimal solution of the LP subproblem

$$\text{maximize } \nabla f(x)^t z,$$

$$\text{subject to } Ax \geq b,$$

with the superscript $t$ denoting the transpose of a vector. Then it seeks a step size $\theta^*$ that is an optimal solution of the one-dimensional subproblem

$$\text{maximize } f(x + \theta d),$$

$$\text{subject to } 0 \leq \theta \leq 1.$$ 

This subproblem is very difficult and usually requires an exhaustive search of the interval in (6). If for all $x \in F^o$ the function in (5) is unimodal over $\theta \in [0, 1]$, then a one-dimensional
search method such as the Method of Golden Sections or the Interval Bisection Method will be convergent to an optimal solution. However, these methods do not generally converge finitely and, therefore, cannot be implemented without a finite stopping rule. Thus, in empirical applications of the F–W Method, two stopping rules are needed, one for the algorithm and the other for its step-size procedure.

Goldstein [3] has used in unconstrained optimization a one-dimensional subproblem for which the step size $\theta^*$ at a point of iteration $x$ that is not a K–T point can be any point $\theta > 0$ such that, given a user-specified constant $b \in (0, 0.5)$,

$$
\theta b \nabla f(x)'d < f(x + \theta d) - f(x) \leq \theta (1 - b) \nabla f(x)'d. (7)
$$

Daniel [1] has shown that for a class of gradient-like feasible direction methods including the F–W Method, if the step size $\theta^*$ is set equal to 1 whenever $f(x + d) \geq f(x) + \theta b \nabla f(x)'d$ and equal to $\theta$ that satisfies Goldstein's rule (given in (7)) otherwise, then $\lim_{n \to \infty} \nabla f(x^n)'d^n = 0$, where $\{x^n\}_{0}$ is a subsequence of iteration points and $\{d^n\}_{0}$ is the corresponding subsequence of direction vectors. For the F–W Method, $\nabla f(x)'d$ is continuous on $F_0$ and, if $\nabla f(x^*)'d^* = 0$ for some $x^* \in F_0$, then $x^*$ is a K–T point. Consequently, modified with Daniel's one-dimensional subproblem, the F–W Method would be convergent to a K–T point. A finite search procedure for this subproblem will be given in the next section.

In order for the algorithm to be computationally implementable, a stopping rule is needed. We consider the possibility of termination at a point of iteration $x$, if $\nabla f(x)'d < \epsilon$, with $\epsilon > 0$ being a user-specified constant. It will be shown that this stopping rule guarantees finite convergence to a point arbitrarily close (depending on the magnitude of the user-specified constant $\epsilon$) to a K–T point. Furthermore, it will be shown that, if the objective function $f$ is concave, then its terminal value will be within $\epsilon$ of the optimal objective function value.

Thus, the F–W Method, embedded with the step-size procedure and the stopping rule, is computationally implementable.

2. Modification of the Algorithm

The algorithm below is the F–W Method, embedded with the step-size procedure and the stopping rule to be discussed.

Algorithm

Step 0. Specify two constants $b \in (0, 0.5)$ and $\epsilon > 0$ and a feasible solution vector $x^{(0)}$. Let $x^{(n)}$ denote the point generated at iteration $n = 1, 2, \ldots$. Set $n = 0$.

Step 1. Let $x = x^{(n)}$ and determine an optimal solution $z^*$ of the LP subproblem defined by (3) and (4). Then compute the direction vector $d = z^* - x$.

Step 3. (Stopping Rule) If $\nabla f(x)'d < \epsilon$, stop ($x$ is sufficiently close to a K–T point).
Step 4. (Step-size Procedure) Let \( x^{(0)} = x + \theta d \). Determine the step size \( \theta^* \) via the following procedure:

A. If \( f(x(1)) \geq f(x) + b\nabla f(x)d \), set \( \theta^* = 1 \). Otherwise let \( \theta_L = 0 \) and \( \theta_H = 1 \) and go to part B.

B. Let \( \theta = (\theta_L + \theta_H)/2 \). If \( f(x(\theta)) > f(x) + \theta(1 - b) \nabla f(x)d \), replace \( \theta_L \) by \( \theta \) and repeat part B. If \( f(x(\theta)) < f(x) + \theta b \nabla f(x)d \), replace \( \theta_H \) by \( \theta \) and repeat part B. Otherwise set \( \theta^* = \theta \).

Step 4. Set \( x^{(n+1)} = x(\theta^*) \). Replace \( n \) by \( n + 1 \) and go to Step 1.

3. Convergence

It is clear that, if the step-size procedure of the algorithm converges to a point \( \theta^* \), then this point will satisfy Daniel’s rule and, as was pointed out in the Introduction, the algorithm without the stopping rule will be convergent to a K-T point. The step-size procedure will always converge finitely, as is shown below.

**THEOREM 1.** The step-size procedure of the algorithm converges in a finite number of iterations.

**Proof.** Consider Fig. 1.

The lines \( L_0 \) and \( L_{1-b} \) have slopes \( b\nabla f(x)d \) and \( (1 - b) \nabla f(x)d \), respectively. The step size \( \theta^* \) can be any point \( \theta > 0 \) such that \( f(x + \theta d) \in [L_0(\theta), L_{1-b}(\theta)] \). Any point from the intervals \([\theta_0, \theta_1], [\theta_2, \theta_3]\) and \([\theta_4, \theta_5]\) is acceptable for the step size.

If \( f(x(1)) \geq f(x) + b\nabla f(x)d \), then \( \theta^* = 1 \) from part A of the procedure. Suppose this is not the case. Then, at every iteration of the procedure would be determined points \( \theta_L \) and \( \theta_H \) such that \( \theta_L < \theta_H \), \( f(x(\theta_L)) > L_{1-b}(\theta_L) \) and \( f(x(\theta_H)) < L_0(\theta_H) \) (see Fig. 1). Clearly, \( f(x(\theta)) \) is continuously differentiable in \( \theta \). Consequently, if \( f(x(\theta)) \) is to be above the line \( L_{1-b} \) at \( \theta = \theta_L \) and below the line \( L_0 \) at \( \theta = \theta_H \), then it must cross the first line at some point \( \theta_3 > \theta_L \) and the second at some point \( \theta_4 < \theta_H \), in such a way that \( f(x(\theta)) \in [L_0(\theta), L_{1-b}(\theta)] \) for all \( \theta \in [\theta_3, \theta_4] \). Thus, at every iteration of the procedure, the interval

![Figure 1](image-url)
THE F–W METHOD

(\(\theta_L, \theta_U\)) will contain at least one sub-interval such as \([\theta_2, \theta_3]\), every point of which would be acceptable for the step size.

It must be that \(\nabla f(x')d > 0\), since the algorithm would have terminated in Step 2 otherwise. Since \(0 < b < 0.5\), it follows that

\[
\nabla f(x')d > (1 - b) \nabla f(x')d > b\nabla f(x')d > 0.
\]

It can be seen that \(df(x(t))/d\theta = \nabla f(x')d\). Consequently, \(df(x(t))/d\theta > (1 - b) \nabla f(x')d\).

Since \(f(x(\theta))\) is continuously differentiable in \(\theta\), it follows that \(f(x(\theta)) > L_{x(\theta)}(\theta)\) for \(\theta\) in some interval \((0, \theta_0)\) having non-zero length. Therefore, \(\theta_2 > 0\). Furthermore, from (8), \(L_{x(\theta)}(\theta_2) > L_{x(\theta_2)}(\theta_2)\). Consequently, since \(f(x(\theta))\) is continuously differentiable in \(\theta\), it must be that \(\theta_3 > \theta_2\) and, thus, the interval \([\theta_2, \theta_3]\) has non-zero length.

Thus, the interval \((\theta_L, \theta_H)\) at every iteration of the procedure contains at least one non-zero-length subinterval \([\theta_2, \theta_3]\) such that every point in this subinterval is acceptable for the step size. The interval \((\theta_L, \theta_H)\) is bisected at every iteration and, thus, contains the interval \((\theta_L, \theta_H)\) of the next iteration. Consequently, there must exist an interval such as \([\theta_2, \theta_3]\) that is contained by all intervals \((\theta_L, \theta_H)\) from the first iteration of the procedure to convergence. Since \((\theta_2, \theta_3)\) has non-zero length, it follows that a point \(\theta \in [\theta_2, \theta_3]\) must be attained in a finite number of iterations.

Q.E.D.

The properties of the stopping rule are derived next.

**Theorem 2.** The stopping rule of the algorithm guarantees finite convergence to a point arbitrarily close to a K–T point.

**Proof.** It is clear that \(x^{(n)} \in F^0\) for all \(n\). Therefore, if \(\nabla f(x^{(n)})d^{(n)} = 0\) for some \(n\), then \(x^{(n)}\) is a K–T point. Furthermore, \(\nabla f(x')d\) is continuous on \(F^0\). Consequently, if \(\nabla f(x^{(n)})d^{(n)}\) is close to 0, then \(x^{(n)}\) is close to a K–T point. On the other hand, the sequence of iteration points \(\{x^{(n)}\}_0^\infty\) (without the stopping rule) has at least one limit point, and every such point is a K–T point. It can be seen that, if \(x^{(n)}\) is a K–T point, then \(x^{(n)}\) is an optimal solution of the LP subproblem in (3) and (4) and, therefore, \(\nabla f(x^{(n)})d^{(n)} = 0\).

It follows that there exists a subsequence \(\{x^{(n)}\}_0^\infty\) such that \(\lim_{n \to \infty} \nabla f(x^{(n)})d^{(n)} = 0\). Consequently, for any \(\epsilon > 0\), there exists an integer \(N < \infty\) such that, for all \(n > N\) in this subsequence, \(\nabla f(x^{(n)})d^{(n)} < \epsilon\). Q.E.D.

For problems with a concave objective function, \(\epsilon\)-convergence to an optimal solution can be achieved, as is shown below.

**Theorem 3.** Suppose that \(f\) is concave on \(F^0\). Then, if \(x^N\) satisfies the stopping rule, \(f(x^N)\) is within \(\epsilon\) of the optimal objective function value.

**Proof.** Clearly, since \(f\) is assumed concave on \(F^0\),

\[
\nabla f(x^{(n)})d(z - x^{(n)}) \geq f(z) - f(x^{(n)}) \quad \text{for} \quad z \in F^0.
\]

(9)
Hence,
\[
\nabla f(x^{(n)})'(x^* - x^{(n)}) = \max_{x \in F^0} \nabla f(x^{(n)})'(x - x^{(n)}) \\
\geq \max_{x \in F^0} f(x) - f(x^{(n)}) \\
= f(x^*) - f(x^{(n)}),
\]
(10)

where \( x^* \) denotes an optimal solution.

If \( x^N \) satisfies the stopping rule, then
\[
\nabla f(x^N)'(x^* - x^N) < \epsilon
\]
(11)
and, from (10), \( f(x^*) - f(x^N) < \epsilon \).

Q.E.D.

4. Empirical Results

The algorithm was applied to a recurring (weekly) sources and uses of funds problem in banking, with a concave objective function. A description of this problem and observations from more than 1000 applications with real data have appeared in [5]. These observations are summarized below:

(1) An initial feasible point is required in Step 0. It is not important to start with a good point; the algorithm makes giant steps in the first few iterations and soon attains a point that is nearly optimal.

(2) The parameter \( b \) in the ste-size procedure can be set equal to any point in the open interval \((0, 0.5)\). Larger \( b \) values generally mean more iterations for the step-size procedure, but fewer iterations for the algorithm itself. In the banking problem, these two opposite effects seem to cancel out one another. So, the speed of convergence appears independent of the \( b \) value used for \( b \) in the interval \([0.01, 0.49]\). However, for \( b \) values close to 0.5, the convergence time would approach infinity as \( b \) approached 0.5, since then the set of acceptable points for the step size would approach a finite set.

(3) In empirical applications of any iterative procedure, it is advisable to place an upper limit on the number of iterations. This is true, even if the procedure is finite and expected to converge within a few iterations, as is true with our step-size procedure, just in case something goes wrong. For the banking problem, the limit on the number of iterations of the step-size procedure was set equal to 50.

(4) The number of iterations to convergence of the algorithm depends, of course, on the value of the parameter \( \epsilon \) in the stopping rule. For the banking problem, with \( \epsilon = 0.001 \) (thus guaranteeing convergence to an objective function value larger than 99.9% of the optimal solution value), the algorithm usually converged within 10 iterations.
ACKNOWLEDGMENT

This paper is an outgrowth of a Ph.D. dissertation [5], submitted to Yale University in 1973. I am thankful to the members of the dissertation committee, Professors Robert Mifflin (chairman), Martin Shubik and Harvey M. Wagner, for their guidance and assistance.

REFERENCES