Analysis and Implementation of a Dual Algorithm for Constrained Optimization\textsuperscript{1,2}

W. W. Hager\textsuperscript{3}

Communicated by E. Polak

Abstract. This paper analyzes a constrained optimization algorithm that combines an unconstrained minimization scheme like the conjugate gradient method, an augmented Lagrangian, and multiplier updates to obtain global quadratic convergence. Some of the issues that we focus on are the treatment of rigid constraints that must be satisfied during the iterations and techniques for balancing the error associated with constraint violation with the error associated with optimality. A preconditioner is constructed with the property that the rigid constraints are satisfied while ill-conditioning due to penalty terms is alleviated. Various numerical linear algebra techniques required for the efficient implementation of the algorithm are presented, and convergence behavior is illustrated in a series of numerical experiments.

Key Words. Constrained optimization, multiplier methods, preconditioning, global convergence, quadratic convergence.

1. Introduction

We consider optimization problems of the following form:

$$\min f(x), \quad \text{s.t. } h(x) = 0, \quad x \in \Omega, \quad \tag{1}$$

where $x$ is a vector in $\mathbb{R}^n$, $f$ is a real-valued function, $h$ maps $\mathbb{R}^n$ to $\mathbb{R}^m$, and $\Omega \subset \mathbb{R}^n$. The constraint set $\Omega$ contains the explicit constraints; these are the constraint that must be satisfied accurately. The constraint $h(x) = 0$, on the

\textsuperscript{1}This research was supported by the National Science Foundation Grant DMS-89-03226 and by the U.S. Army Research Office Contract DAA03-89-M-0314.

\textsuperscript{2}We thank the referees for their many perceptive comments which led to substantial improvements in the presentation of this paper.

\textsuperscript{3}Professor, Department of Mathematics, University of Florida, Gainesville, Florida.
other hand, will be satisfied approximately in our numerical algorithms, and as the iterations progress, the constraint violation will be reduced. In our numerical experiments, we include all nonlinear constraints in \( h \), while the linear equalities and inequalities are incorporated in the constraint \( x \in \Omega \). More precisely, in our computer code, we assume that \( \Omega \) is given by

\[
\Omega = \{ x: Ax = b, l \leq x \leq u \},
\]

(2)

where \( A \) is a matrix and \( b \) is a vector of compatible dimensions, \( l \) is a vector of lower bounds, and \( u \) is a vector of upper bounds. Of course, any optimization problem constrained by systems of equalities and inequalities can be expressed in the form (1) with \( \Omega \) given by (2).

Our goal is to find a feasible point that satisfies the Kuhn–Tucker condition associated with (1). Let \( L \) denote the Lagrangian defined by

\[
L(\lambda, x) = f(x) + \lambda^T h(x),
\]

where the superscript \( T \) denotes transpose. If \( \Omega \) is given by a system of equalities and inequalities,

\[
\Omega = \{ x \in \mathbb{R}^n: g(x) = 0 \text{ and } G(x) \leq 0 \},
\]

where \( g \) maps \( \mathbb{R}^n \) to \( \mathbb{R}^l \) and \( G \) maps \( \mathbb{R}^n \) to \( \mathbb{R}^k \), then a point \( x \in \Omega \) with \( h(x) = 0 \) satisfies the Kuhn–Tucker condition associated with (1) if there exists a vector \( \lambda \in \mathbb{R}^m \) such that

\[
-\nabla_x L(\lambda, x) \in N_x(\Omega),
\]

where \( \nabla_x \) stands for the gradient with respect to \( x \) and \( N_x(\Omega) \) is the normal cone defined by

\[
N_x(\Omega) = \{ \mu^T \nabla g(x) + v^T \nabla G(x): v \geq 0, v_i = 0 \text{ if } g_i(x) < 0, \mu \text{ arbitrary} \}.
\]

Given \( x \in \Omega \) and \( \lambda \in \mathbb{R}^m \), let \( E \) denote the error expression defined by

\[
E(\lambda, x) = K(\lambda, x) + C(x),
\]

where \( C(x) = |h(x)| \) measures the constraint violation in some norm \( |\cdot| \) and \( K \) measures the error in the Kuhn–Tucker condition,

\[
K(\lambda, x) = \text{distance}\{-\nabla_x L(\lambda, x), N_x(\Omega)\}
\]

\[
= \text{minimum}\{|\nabla_x L(\lambda, x) + y|: y \in N_x(\Omega)\}.
\]

For convenience, we use the Euclidean norm throughout this paper, although any norm can be employed.