Measures of Optimality for Constrained Optimization*

Elizabeth D. Dolan,† Jorge J. Moré,§ and Todd S. Munson§

Draft: April 16, 2002

Given an optimization problem defined by an objective function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and constraints $c : \mathbb{R}^n \mapsto \mathbb{R}^m$, we define measures of optimality for the general optimization problem

$$\min \{ f(x) : l \le c(x) \le u \} . \tag{1}$$

Our aim is to benchmark the accuracy achieved by optimization algorithms. A secondary aim is to test identification functions.

1 Background

Given $x \in \mathbb{R}^n$, we define measures of optimality in terms of the set of ϵ-active constraints. An ϵ-active constraint is defined in terms of the distance to the boundary of x. Given x, y in $\mathbb{R}^n$, we define a distance between vectors by

$$\delta_k[x,y] = \min \left\{ |x_k - y_k|, \frac{|x_k - y_k|}{\min\{|x_k|, |y_k|\}} \right\}$$

if $\min\{|x_k|, |y_k|\} \neq 0$; otherwise, $\delta_k[x,y] = |x_k - y_k|$. The set of ϵ-active constraints at x is then

$$\mathcal{A}_\epsilon(x) = \{ k : \min\{\delta_k[c(x),l], \delta_k[c(x),u]\} \le \epsilon \} .$$

In general ϵ is related to the expected accuracy of the optimization algorithm since the set $\mathcal{A}_\epsilon(x)$ contains all constraints that are nearly active as measured by ϵ. In all cases, we must expect that

$$l_k \le c_k(x) \le u_k \quad \text{or} \quad \min\{\delta_k[c(x),l], \delta_k[c(x),u]\} \le \epsilon, \qquad 1 \le k \le n.$$

*This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38, and by the National Science Foundation (Information Technology Research) grant CCR-0082807.

†Department of Industrial Engineering & Management Sciences, Northwestern University, and Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439 (`dolan@mcs.anl.gov`).

§Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439 ({`more,munson`}`@mcs.anl.gov`).

2 Multipliers

We measure optimality by computing multipliers explicitly. We need the matrix of constraint gradients $C(x)$ for the constraints that are ϵ-active, that is,

$$C(x) = \left(\nabla c_k(x) : k \in \mathcal{A}_\epsilon\right).$$

Similarly, we need the critical cone defined by

$$S_\epsilon = \left\{ v : \begin{array}{lll} v_k \text{ free} & \text{if} & \delta_k[c(x), l] \le \epsilon,\ \delta_k[c(x), u] \le \epsilon \\ v_k \ge 0 & \text{if} & \delta_k[c(x), l] \le \epsilon \\ v_k \le 0 & \text{if} & \delta_k[c(x), u] \le \epsilon \end{array} \right.$$

We now determine multipliers via the bound-constrained least squares problem

$$\min \left\{ \|\nabla f(x) - C(x)v\| : v \in S_\epsilon \right\}. \tag{2}$$

If $\lambda(x)$ is a solution of (2), then

$$\nu_s(x) = \|\nabla f(x) - C(x)\lambda(x)\| \tag{3}$$

is an absolute measure of optimality. In the special case where S_ϵ is empty (the problem may be unconstrained) we set $\nu_s(x) = \|\nabla f(x)\|$.

The AMPL/GAMS facilities allow us to formulate the computation of ν_s in different ways. Since there are no solvers that deal with the structure of the least squares problem (2), we prefer to compute the multipliers via the problem

$$\min \left\{ \tfrac{1}{2}\|y\|^2 : y = \nabla f(x) - C(x)v,\ v \in S_\epsilon \right\}. \tag{4}$$

In this formulation, $\nu_s(x) = \|y\|$.

We verify that the computation of the multipliers via (2) or (4) is accurate by computing the projected gradient. For this computation, the gradient is

$$r = C(x)^T \left(C(x)v - \nabla f(x)\right),$$

so that the projected gradient is

$$\hat{r}_k = \left\{ \begin{array}{lll} r_k & \text{if} & \delta_k[c(x), l] \le \epsilon,\ \delta_k[c(x), u] \le \epsilon \\ \min(r_k, 0) & \text{if} & \delta_k[c(x), l] \le \epsilon \\ \max(r_k, 0) & \text{if} & \delta_k[c(x), u] \le \epsilon \end{array} \right. \tag{5}$$

3 Feasibility and Complementarity

We measure feasibility in terms of relative and absolute distances to the boundary with the function

$$\nu_f(x) = \max\{\mu_k(x) : 1 \le k \le n\} \tag{6}$$

where

$$\mu_k(x) = \begin{cases} 0 & \text{if} \quad l_k \le c_k(x) \le u_k, \\ \delta_k[c(x), l] & \text{if} \quad c_k(x) \le u_k, \qquad \text{and} \\ \delta_k[c(x), u] & \text{if} \quad l_k \le c_k(x). \end{cases}$$

If $\nu_f(x) \le \epsilon$ then the problem is ϵ- feasible.

An advantage of computing the multipliers by either (2) or (4) is that all the multipliers of the ϵ-active constraints have the proper sign. We define

$$\nu_c(x) = \max\left\{\min(\delta_k[c(x), l], \delta_k[c(x), u)]) : k \in \mathcal{A}_\epsilon(x)\right\}, \tag{7}$$

as a measure of complementarity. Note that $\epsilon \mapsto \nu_c(x)$ is non-decreasing.

4 Optimality

Given $x \in \mathbb{R}^n$ we have outlined three measures of optimality for the general optimization problem (1). A benchmark solver should provide these measures.

We have the *distance to feasibility* $\nu_f(x)$ defined by (6). Note that this measure combines relative and absolute distances to the boundary.

We also have the *distance to complementarity* $\nu_c(x)$ defined by (7). This measure combines relative and absolute distances to the boundary for those constraints that are considered to be ϵ-active. Note that our definition guarantees that $\nu_c(x) \le \epsilon$.

The final measure of optimality is the *distance to a Kuhn-Tucker point* defined by (3). Since this measure depends on the scaling of the gradient, we also want to know the relative measure,

$$\nu_{s,r}(x) = \frac{\|\nabla f(x) - C(x)\lambda(x)\|}{\|\nabla f(x)\|},$$

of optimality. Finally, we provide the *absolute multiplier accuracy* and the *relative multiplier accuracy*,

$$\nu_m(x) = \|\hat{r}\|, \qquad \nu_{m,r}(x) = \frac{\|\hat{r}\|}{\|\nabla f(x)\|},$$

where $\hat{r}$ is the projected gradient defined by (5).

feasibility	$\nu_s(x)$
complementarity	$\nu_c(x)$
optimality (absolute)	$\nu_s(x)$
optimality (relative)	$\nu_{s,r}(x)$
multiplier (absolute)	$\nu_m(x)$
multiplier (relative)	$\nu_{m,r}(x)$

Table 1: Measures of optimality

Table 1 summarizes all the measures of optimality for an optimization problem. We consider the multipliers to be sufficiently accurate if

$$\min\{\nu_m(x), \nu_{m,r}(x)\} \le \sigma_1 \min\{\nu_c(x), \nu_{s,r}(x), \sigma_2\}$$

for some tolerances $\sigma_1,\ \sigma_2$ in $(0,1)$. For example, $\sigma_1 = 0.1$ and $\sigma_2 = 10^{-3}$.

5 Multipliers via Steepest Descent

An alternate definition of multipliers is obtained by considering the steepest descent direction with respect to ϵ-feasible directions. We define the set of ϵ-feasible direction as the set $F_\epsilon(x)$ of all $w \in \mathbb{R}^n$ that satisfy the following conditions

$$\begin{array}{ll} \langle \nabla c_k(x), w \rangle \geq 0, & \delta_k[c(x), l] \leq \epsilon, \\ \langle \nabla c_k(x), w \rangle \leq 0, & \delta_k[c(x), u] \leq \epsilon, \end{array}$$

Note that if $\delta_k[c(x), l] \leq \epsilon$ and $\delta_k[c(x), u] \leq \epsilon$, then we require that $\langle \nabla c_k(x), w \rangle = 0$. This happens if $l_k = u_k$, but may also happen if l_k and u_k are close.

We can now define the steepest descent direction relative to the set $\mathcal{A}_\epsilon$ as the solution of the optimization problem

$$\min \left\{ \langle \nabla f(x), w \rangle : \|w\| \leq 1,\ w \in F_\epsilon \right\}.$$

If w is the solution to this problem, then

$$\nu_s(x) = -\langle \nabla f(x), w \rangle,$$

is equivalent to the definition of ν_s via (3). In general we can only say that

$$0 \leq \nu_s(x) \leq \|\nabla f(x)\|,$$

and that the lower bound is achieved for large ϵ in most cases. Also note that the function $\epsilon \mapsto \nu_s(x)$ is non-increasing.

6 Theory

We want to justify these measures by showing that if

$$\nu(x) = \max(\nu_f(x), \nu_s(x), \nu_c(x))$$

then $\nu(x) \leq \epsilon$ if and only x is an approximate solution of the constrained optimization problem (1). We also need to study the choice of ϵ. We should choose ϵ so that $\nu(x)$ above is (nearly) minimal.

Theorem 1 *If the MFCQ is satisfied at x^*, then $x \mapsto \nu(x)$ is lower semicontinuous at x^*.*