Distance-to-Solution Estimates for Optimization Problems with Constraints in Standard Form

Philip E. Gill* Vyacheslav Kungurtsev† Daniel P. Robinson‡

UCSD Center for Computational Mathematics
Technical Report CCoM-16-1
April 20, 2016

Abstract

An important tool in the formulation and analysis of algorithms for constrained optimization is a quantity that provides a practical estimate of the distance to the set of primal-dual solutions. Such "distance-to-solution estimates" may be used to identify the inequality constraints satisfied with equality at a solution, and to formulate conditions used to terminate a sequence of solution estimates. This note concerns the properties of a particular distance-to-solution estimate for optimization problems with constraints written in so-called "standard form", which is a commonly-used approach for formulating constraints with a mixture of equality and inequality constraints.

Key words. Nonlinear programming, nonlinear optimization, inequality constraints, primal-dual methods, second-order optimality.

1. Introduction

An important tool in the formulation and analysis of algorithms for constrained optimization is a quantity that provides a practical estimate of the distance to the set of primal-dual solutions (see Facchinei, Fischer and Kanzow [1], Hager and Gowda [3], Wright [4,5], and Gill, Kungurtsev and Robinson [2]). Such "distance-to-solution estimates" may be used to identify the inequality constraints satisfied with equality at a solution, and to formulate conditions used to terminate a sequence of solution estimates. This note concerns the properties of a particular distance-to-solution estimate for optimization problems with constraints written in so-called "standard form", which is a commonly-used approach for formulating constraints with a mixture of equality and inequality constraints. Any optimization problem with smooth problem functions and a mixture of equality and inequality constraints may be written in the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\;\; f(x) \quad \text{subject to} \quad c(x) = 0, \quad x \ge 0, \tag{NP}$$

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). Research supported by National Science Foundation grants DMS-0915220 and DMS-1318480.

†Agent Technology Center, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. (vyacheslav.kungurtsev@fel.cvut.cz) Research supported by the European social fund within the framework of realizing the project "Support of inter-sectoral mobility and quality enhancement of research teams at the Czech Technical University in Prague", CZ.1.07/2.3.00/30.0034.

‡Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682 (daniel.p.robinson@jhu.edu). Research supported by National Science Foundation grant DMS-1217153.

where $c\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ are smooth vector- and scalar-valued functions. Throughout it is assumed that the problem functions c and f are twice Lipschitz-continuously differentiable.

A vector x^* is a first-order KKT point for problem (NP) if there exists a dual vector y^* such that

$$\begin{aligned} c(x^*) &= 0, & x^* &\geq 0, \\ x^* \cdot \big(g(x^*) - J(x^*)^T y^*\big) &= 0, & g(x^*) - J(x^*)^T y^* &\geq 0. \end{aligned} \tag{1.1}$$

where $g(x)$ denotes $\nabla f(x)$, the gradient of f at x, and $J(x)$ denotes the $m \times n$ constraint Jacobian matrix, which has ith row $\nabla c_i(x)^T$, the gradient of the ith constraint function c_i at x. The KKT conditions (1.1) may be written in the equivalent form $r(x^*, y^*) = 0$, where

$$r(x,y) = \left\| \big(c(x), \min\big(x, g(x) - J(x)^T y\big)\big) \right\|. \tag{1.2}$$

Any (x^*, y^*) satisfying (1.1) or, equivalently, $r(x^*, y^*) = 0$, is called a first-order KKT pair. For arbitrary vectors x and y of appropriate dimension, the scalar $r(x,y)$ provides a practical estimate of the distance of (x,y) to a first-order KKT pair of problem (NP). In general, the Lagrange multiplier associated with a first-order KKT point is not unique, and the set of Lagrange multiplier vectors is given by

$$\mathcal{Y}(x^*) = \{ y \in \mathbb{R}^m : (x^*, y) \ \text{ satisfies } \ r(x^*, y) = 0 \}. \tag{1.3}$$

Second-order optimality conditions involve the properties of the active set of nonnegativity constraints $\mathcal{A}(x) = \{ i : [\,x\,]_i = 0 \}$, and the sets of strongly-active variables $\mathcal{A}_+(x,y)$ and weakly-active variables $\mathcal{A}_0(x,y)$, with

$$\begin{aligned} \mathcal{A}_+(x,y) &= \{ i \in \mathcal{A}(x) : [\,g(x) - J(x)^T y\,]_i > 0 \}, \\ \mathcal{A}_0(x,y) &= \{ i \in \mathcal{A}(x) : [\,g(x) - J(x)^T y\,]_i = 0 \}. \end{aligned} \tag{1.4}$$

where $\mathcal{A}(x)$ denotes the active set of nonnegativity constraints at x, i.e., $\mathcal{A}(x) = \{ i : [\,x\,]_i = 0 \}$. A primal-dual pair (x^*, y^*) satisfies the second-order sufficient optimality conditions (SOSC) for problem (NP) if it is a first-order KKT pair (i.e., $r(x^*, y^*) = 0$) and $p^T H(x^*, y^*) p > 0$ for all $p \in \mathcal{C}(x^*, y^*) \setminus \{0\}$, where $\mathcal{C}(x^*, y^*)$ is the critical cone

$$\mathcal{C}(x^*, y^*) = \operatorname{null}\big(J(x^*)\big) \cap \{ p : p_i = 0 \ \text{ for } \ i \in \mathcal{A}_+(x^*, y^*), \ p_i \geq 0 \ \text{ for } \ i \in \mathcal{A}_0(x^*, y^*) \, \}.$$

The results require the following assumption.

Assumption 1.1. *If (x^*, y^*) is a first-order KKT pair, then*

(i) *there exists a compact set $\Lambda(x^*) \subseteq \mathcal{Y}(x^*)$ such that y^* belongs to the (nonempty) interior of $\Lambda(x^*)$ relative to $\mathcal{Y}(x^*)$; and*

(ii) *(x^*, y) satisfies the second-order sufficient conditions for every $y \in \Lambda(x^*)$.*

The existence of the compact set $\Lambda(x^*)$ of Assumption 1.1 guarantees that the closest point in $\mathcal{Y}(x^*)$ to every element y_k of a sequence $\{y_k\}$ satisfying $\lim_{k\to\infty} y_k = y^*$ is also in $\Lambda(x^*)$ for k sufficiently large. This is equivalent to there being a set $\mathcal{K}$, open relative to $\mathcal{Y}(x^*)$, such that $y^* \in \mathcal{K} \subset \Lambda(x^*)$. This implies that the affine hulls of $\Lambda(x^*)$ and $\mathcal{Y}(x^*)$ are identical, with y^* in the relative interior of $\Lambda(x^*)$. Note that the set of multipliers $\mathcal{Y}(x^*)$ need not be bounded. The second-order sufficient conditions need hold only for multipliers in a compact subset of $\mathcal{Y}(x^*)$.

2. Distance to the solution of a problem in standard form

We start by defining a quantity that measures the distance of a primal-dual point (x, y) to the primal-dual solution set of problem (NP). Let (x^*, y^*) denote a primal-dual pair satisfying the second-order sufficient optimality conditions. For any given y, the compactness of $\Lambda(x^*)$ implies the existence of a vector $y_P^*(y) \in \Lambda(x^*)$ that minimizes the distance from y to the set $\Lambda(x^*)$, i.e.,

$$y_P^*(y) \in \operatorname*{Argmin}_{\bar{y}\in\Lambda(x^*)} \|y - \bar{y}\|. \tag{2.1}$$

The existence of a vector $y_P^*(y)$ implies that the distance $\delta(x, y)$ of any primal-dual point (x, y) to the primal-dual solution set $\mathcal{V}(x^*) = \{x^*\} \times \Lambda(x^*)$ associated with x^*, may be written in the form

$$\delta(x, y) = \min_{(\bar{x},\bar{y})\in\mathcal{V}(x^*)} \|(x - \bar{x}, y - \bar{y})\| = \|(x - x^*, y - y_P^*(y))\|. \tag{2.2}$$

The results of this section show that the proximity measure $r(x, y)$ of (1.2) may be used as a surrogate for $\delta(x, y)$ near (x^*, y^*). The results involve the relationship between the proximity measure $r(x, y)$, and the quantities $\eta(x, y)$ and $\bar{\eta}(x, y)$ defined by Wright [5] (and also defined below). Throughout the discussion, the scaled closed interval $[\delta\,\alpha_\ell, \delta\,\alpha_u]$ defined in terms of the positive scale factor δ and positive scalars α_ℓ and α_u, will denoted by $[\alpha_\ell, \alpha_u] \cdot \delta$.

2.1. Distance to the solution estimates in inequality-constraint form

The main result (Theorem 2.1 below) relies on several results of Wright [5], which concern an optimization problem with all inequality constraints. The all-inequality form of problem (NP) is

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} \;\; f(x) \quad \text{subject to} \quad c(x) \geq 0, \quad -c(x) \geq 0, \quad x \geq 0. \tag{2.3}$$

Given multipliers y for problem (NP), the multipliers for the nonnegativity constraints $x \geq 0$ are $g(x) - J(x)^T y$ and are denoted by $z(x, y)$.

Consider the primal-dual solution set $\mathcal{V}_z(x^*)$ for problem (2.3). It follows that $\mathcal{V}_z(x^*) = \mathcal{V}(x^*) \times \mathcal{Z}(x^*)$, where

$$\mathcal{V}(x^*) = \{x^*\} \times \Lambda(x^*) \;\text{ and }\; \mathcal{Z}(x^*) = \{z : g(x^*) - J(x^*)^T y, \;\text{ for some }\; y \in \Lambda(x^*)\}.$$

The distance to optimality for the problem (2.3) is then

$$\begin{aligned}\operatorname{dist}\big((x, y, z), \mathcal{V}_z(x^*)\big) &= \min_{(\bar{x},\bar{y},\bar{z})\in\mathcal{V}_z(x^*)} \|(x - \bar{x}, y - \bar{y}, z - \bar{z})\| \\ &= \min_{(\bar{x},\bar{y})\in\mathcal{V}(x^*)} \big\|\big(x - \bar{x}, y - \bar{y}, z(x, y) - (g(\bar{x}) - J(\bar{x})^T\bar{y})\big)\big\|.\end{aligned}$$

Lemma 2.1. *If* $\operatorname{dist}\big((x, y, z), \mathcal{V}_z(x^*)\big)$ *denotes the distance to optimality for the problem* (NP) *written in all-inequality form, then* $\delta(x, y) = \Theta\big(\operatorname{dist}((x, y, z), \mathcal{V}_z(x^*))\big)$.

Proof. Let $y_P^*(y)$ denote the vector that minimizes the distance from y to the compact set $\Lambda(x^*)$ (see (2.1)). Consider the quantity

$$\underline{\delta}(x, y) = \big\|\big(x - x^*,\, y - y_P^*(y),\, z(x, y) - z(x^*, y_P^*(y))\big)\big\|.$$

The components of the vector $\big(x - x^*,\, y - y^*_P(y)\big)$ used to define $\delta(x, y)$ form the first $n + m$ components of $\underline{\delta}(x, y)$, which implies that $\delta(x, y) \le \underline{\delta}(x, y)$. For the upper bound, the Lipschitz continuity of J and g, together with the boundedness of $y^*_P(y)$ and $J(x)$ imply that

$$\begin{aligned} \|z(x,y) - z(x^*, y^*_P(y))\| &= \|g(x) - J(x)^T y - g(x^*) + J(x^*)^T y^*_P(y)\| \\ &\le L_g\|x - x^*\| + \|J(x)^T(y - y^*_P(y))\| + \|(J(x) - J(x^*))^T y^*_P(y)\| \\ &\le L_g\|x - x^*\| + C_J\|y - y^*_P(y)\| + L_2 C_y\|x - x^*\| \\ &\le C_a\delta(x, y). \end{aligned} \tag{2.4}$$

It follows that $\underline{\delta}(x, y) \le \delta(x, y) + \|z(x, y) - z(x^*, y^*_P(y))\| \le (1 + C_a)\delta(x, y)$, which implies that $\underline{\delta}(x, y) = \Theta\big(\delta(x, y)\big)$, and, equivalently, $\delta(x, y) = \Theta\big(\underline{\delta}(x, y)\big)$.

The proof is complete if it can be shown that $\delta(x, y) = \Theta\big(\operatorname{dist}((x, y, z), \mathcal{V}_z(x^*))\big)$. The definitions of $\operatorname{dist}((x, y, z), \mathcal{V}_z)$ and $\underline{\delta}(x, y)$ imply that $\operatorname{dist}((x, y, z), \mathcal{V}_z(x^*)) \le \underline{\delta}(x, y)$. Moreover, $\delta(x, y) = \operatorname{dist}((x, y), \mathcal{V}(x^*)) \le \operatorname{dist}((x, y, z), \mathcal{V}_z(x^*))$ because there is no third component in the definition of $\delta(x, y)$. As $\delta(x, y) = \Theta\big(\underline{\delta}(x, y)\big)$, it must hold that $\delta(x, y) = \Theta\big(\operatorname{dist}((x, y, z), \mathcal{V}_z(x^*))\big)$, as required. ∎

Let $\eta(x, y)$ be the practical estimate of $\operatorname{dist}\big((x, y, z), \mathcal{V}_z(x^*)\big)$ given by

$$\eta(x, y) = \|\big(v_1(x, y), v_2(x, y), v_3(x, y), v_4(x, y)\big)\|_1,$$

where $v_1 = g(x) - J(x)^T y - z(x, y)$, $v_2 = \min(x, z(x, y))$, $v_3 = \min(c(x), \max(y, 0))$, and $v_4 = \min(-c(x), \max(-y, 0))$. Wright [5, Theorem 3.2] shows that under Assumption 1.1, it holds that

$$\eta(x, y) \in [1/\kappa, \kappa] \cdot \operatorname{dist}\big((x, y, z), \mathcal{V}_z(x^*)\big)$$

for all (x, y) sufficiently close to (x^*, y^*).

Lemma 2.2. *Consider the function $\eta(x, y) = \|(v_1, v_2, v_3, v_4)\|_1$, where $v_1 = g(x) - J(x)^T y - z(x, y)$, $v_2 = \min(x, z(x, y))$, $v_3 = \min(c(x), \max(y, 0))$, and $v_4 = \min(-c(x), \max(-y, 0))$. The quantity $\eta(x, y)$ defines a measure of the quality of (x, y) as an approximate solution of problem* (NP) *defined in all-inequality form and satisfies $r(x, y) = \Theta\big(\eta(x, y)\big)$.*

Proof. It will be established that $\eta(x, y) = \Theta\big(r(x, y)\big)$. The vector v_1 is zero by definition. The vector v_2 is $\min(x, g(x) - J(x)^T y)$, which is the second part of $r(x, y)$.

If $c_i(x) < 0$ and $y_i \ge 0$ then $\min(c_i(x), \max(y_i, 0)) = c_i(x)$ and $\min(-c_i(x), \max(-y_i, 0)) = 0$. If $c_i(x) < 0$ and $y_i \le 0$ then $\min(c_i(x), \max(y_i, 0)) = c_i(x)$ and $\min(-c_i(x), \max(-y_i, 0)) = \min(|c_i(x)|, |y_i|)$. If $c_i(x) > 0$ and $y_i \ge 0$ then $\min(c_i(x), \max(y_i, 0)) = \min(|c_i(x)|, |y_i|)$ and $\min(-c_i(x), \max(-y_i, 0)) = -c_i(x)$. If $c_i(x) > 0$ and $y_i \le 0$ then $\min(c_i(x), \max(y_i, 0)) = 0$ and $\min(-c_i(x), \max(-y_i, 0)) = -c_i(x)$.

It follows that for every i, one or the other of the vectors v_3 or v_4 has a component equal to $|c_i(x)|$ and hence $\eta(x, y) \ge r(x, y)$. In addition, v_3 or v_4 may have a term that is $\min(|c_i(x)|, |y_i|) \le |c_i(x)|$, and so $\eta(x, y) \le 2r(x, y)$. It follows that $\eta(x, y) = \Theta\big(r(x, y)\big)$, as required. ∎

Theorem 2.1. **([5, Theorem 3.2])** *There exists a constant positive scalar $\kappa \equiv \kappa(\Lambda(x^*))$ such that $r(x, y) \in \big[\,\delta(x, y)/\kappa,\ \delta(x, y)\kappa\,\big]$ for all (x, y) sufficiently close to (x^*, y^*).*

Proof. Under the assumptions used here, the result follows from Theorem 3.2 of Wright [5], where Lemmas 2.2 and 2.1 are used to establish that the exact and estimated distance of (x, y) to the primal-dual solution set used in [5] are equivalent (up to a scalar multiple) to the values $\delta(x, y)$ and $r(x, y)$ given here. ∎

2.2. Equality-constraint form

An important property of any effective active-set method for constraint optimization is the ability to identify the active set associated with the inequality constraints satisfied with equality at a solution x^*. Once this set has been identified, the active-set method either implicitly or explicitly solves a problem of the form

$$\underset{x}{\text{minimize}}\; f(x) \quad \text{subject to} \quad c(x) = 0, \quad \text{and} \quad x_{\mathcal{A}} = E_{\mathcal{A}}^T x = 0, \tag{2.5}$$

where $E_{\mathcal{A}}^T$ is the matrix of gradients of the nonnegativity constraints that are active at x^*, and $x_{\mathcal{A}}$ denotes the vector of components of x with indices in $\mathcal{A}(x^*)$.

Any primal-dual solution (x^*, y^*) of problem (NP) must satisfy the SOSC for (2.5) because the conditions for problem (NP) imply that $p^T H(x^*, y^*)p > 0$ for all p such that $J(x^*)p = 0$ and $p_i = 0$ for every $i \in \mathcal{A}(x^*)$. The primal-dual solution set $\mathcal{U}_z(x^*)$ for problem (2.5) has the form $\mathcal{U}_z(x^*) = \mathcal{U}(x^*) \times \mathcal{Z}(x^*)$, where

$$\mathcal{U}(x^*) = \{x^*\} \times \Lambda(x^*) \;\text{ and }\; \mathcal{Z}(x^*) = \{z_{\mathcal{A}} : z_{\mathcal{A}} = [\, g(x^*) - J(x^*)^T y\,]_{\mathcal{A}}, \;\text{ for some }\; y \in \Lambda(x^*)\},$$

with $[\, g(x) - J(x)^T y\,]_{\mathcal{A}}$ the vector of components of $g(x) - J(x)^T y$ with indices in $\mathcal{A}(x^*)$. Let y and $z_{\mathcal{A}}$ denote estimates of the multipliers for the constraints $c(x) = 0$ and $E_{\mathcal{A}}^T x = 0$. Let $\bar{\delta}(x, y, z_{\mathcal{A}})$ be the distance of $(x, y, z_{\mathcal{A}})$ to a solution of (2.5), i.e.,

$$\begin{aligned}
\operatorname{dist}(x, y, z_{\mathcal{A}}, \mathcal{U}_z(x^*)) &= \min_{(\bar{x}, \bar{y}, \bar{z}_{\mathcal{A}}) \in \mathcal{U}_z(x^*)} \|(x - \bar{x}, y - \bar{y}, z_{\mathcal{A}} - \bar{z}_{\mathcal{A}})\| \\
&= \min_{(\bar{x}, \bar{y}) \in \mathcal{U}(x^*)} \|(x - \bar{x}, y - \bar{y}, [\, g(x) - J(x)^T y - (g(\bar{x}) - J(\bar{x})^T \bar{y})\,]_{\mathcal{A}})\| \\
&= \min_{\bar{y} \in \Lambda(x^*)} \|(x - x^*, y - \bar{y}, [\, g(x) - J(x)^T y - (g(x^*) - J(x^*)^T \bar{y})\,]_{\mathcal{A}})\|,
\end{aligned}$$

where $\Lambda(x^*)$ is the compact subset of the set of optimal multipliers corresponding to x^* for problem (NP).

Let $\widetilde{\mu}(x, y, z_{\mathcal{A}})$ be the estimate of $\operatorname{dist}(x, y, z_{\mathcal{A}}, \mathcal{U}_z(x^*))$ given by

$$\widetilde{\mu}(x, y, z_{\mathcal{A}}) = \left\| \begin{pmatrix} g(x) - J(x)^T y - E_{\mathcal{A}} z_{\mathcal{A}} \\ c(x) \\ x_{\mathcal{A}} \end{pmatrix} \right\| = \left\| \begin{pmatrix} [\, g(x) - J(x)^T y\,]_{\mathcal{A}} - z_{\mathcal{A}} \\ [\, g(x) - J(x)^T y\,]_{\mathcal{F}} \\ c(x) \\ x_{\mathcal{A}} \end{pmatrix} \right\|_1, \tag{2.6}$$

where $[\, g(x) - J(x)^T y\,]_{\mathcal{F}}$ denotes the vector of components of $g(x) - J(x)^T y$ with indices $i \notin \mathcal{A}(x^*)$. Wright [5] uses $\bar{\eta}(x, y, z_{\mathcal{A}})$ to denote the quantity $\widetilde{\mu}(x, y, z_{\mathcal{A}})$ and shows that for all (x, y) sufficiently close to (x^*, y^*), the estimate $\widetilde{\mu}(x, y, z_{\mathcal{A}})$ satisfies

$$\widetilde{\mu}(x, y, z_{\mathcal{A}}) \in [1/\kappa, \kappa] \cdot \operatorname{dist}(x, y, z_{\mathcal{A}}, \mathcal{U}_z(x^*)), \tag{2.7}$$

where $\kappa = \kappa(\mathcal{U}_z(x^*))$ is a constant.

Lemma 2.3. *For all (x, y) sufficiently close to (x^*, y^*), the estimate $\widetilde{\mu}(x, y, z_{\mathcal{A}}) = \|(g(x) - J(x)^T y - E_{\mathcal{A}} z_{\mathcal{A}}, c(x), x_{\mathcal{A}})\|_1$ satisfies $\widetilde{\mu}(x, y, z_{\mathcal{A}}) = O(\delta(x, y))$.*

Proof. For all (x, y) sufficiently close to (x^*, y^*), the definition of $\operatorname{dist}(x, y, z_{\mathcal{A}}, \mathcal{U}_z(x^*))$ and the Lipschitz continuity of g and J imply that

$$\begin{aligned}\operatorname{dist}(x, y, z_{\mathcal{A}}, \mathcal{U}_z(x^*)) &\leq \delta(x, y) + \|[\, g(x) - J(x)^T y - (g(x^*) - J(x^*)^T y^*_P(y))\,]_{\mathcal{A}}\| \\ &\leq \delta(x, y) + \|g(x) - J(x)^T y - (g(x^*) - J(x^*)^T y^*_P(y))\| \\ &\leq \delta(x, y) + C_a \delta(x, y),\end{aligned}$$

for some bounded constant C_a (cf. (2.4)). The result now follows from (2.7). ∎

References

[1] F. Facchinei, A. Fischer, and C. Kanzow. On the accurate identification of active constraints. *SIAM J. Optim.*, 9(1):14–32, 1998. 1

[2] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: Superlinear convergence. Center for Computational Mathematics Report CCoM 14-01, University of California, San Diego, 2014. Revised June 2015. 1

[3] W. W. Hager and M. S. Gowda. Stability in the presence of degeneracy and error estimation. *Math. Program.*, 85(1, Ser. A):181–192, 1999. 1

[4] S. J. Wright. Modifying SQP for degenerate problems. *SIAM J. Optim.*, 13(2):470–497, 2002. 1

[5] S. J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM J. Optim.*, 15(3):673–696, 2005. 1, 3, 4, 5