Chapter 5

Inequality-Constrained Optimization

5.1 Introduction

The constraints under which a design problem is formulated involve, more often than not, inequalities constraints, in addition to equalities. In fact, inequality constraints arise naturally in design because the resources available to accomplish a design job are *finite*. For example, a designed object must: fit into a designated region; be realizable within a given budget; and be delivered by a certain date.

In this chapter we address two issues around inequality-constrained problems: the *normality conditions* and the *methods of solution*. As to the former, we will not dwell into their rigorous derivation, which are elusive to a simple analysis with the tools of linear algebra; rather, we will introduce the first-order normality conditions without derivation, and illustrate their validity with examples. The second-order normality conditions will be derived using a pragmatic approach, more so than a mathematical formulation.

The reason why linear algebra is no longer sufficient to derive the normality conditions of inequality-constrained problems lies in the nature of inequalities, which define regions of $\mathbb{R}^n$ that are neither vector spaces nor manifolds, as we encountered when studying equality-constrained problems. Now we will speak, more generally, of the *feasible region*, which can have sharp edges and vertices, not occurring in manifolds. For this reason, a simple transformation of the form $\mathbf{x}(\mathbf{u})$ is not sufficient, in general, to guarantee the fulfilment of the inequality constraints.

Regarding the methods of solution of inequality-constrained design problems, we will proceed as usual: we will transform these problems into equality-constrained problems, which we know how to handle. The outcome is that we can apply the whole apparatus developed for equality-constrained problems when dealing with their inequality-constrained counterparts. That is, the ODA package can be applied successfully to the solution of inequality-constrained problems as well, if with the applicable provisions, as described here.

5.2 The Karush-Kuhn-Tucker Conditions

The first-order normality conditions of equality-constrained problems are classical results, first proposed by Joseph Louis de Lagrange, brilliant mathematician born in Turin in 1736 and dead in Paris in 1813. Lagrange founded in Turin a society that would become the Academy of Sciences; then, Lagrange went to Berlin, to the Academy of Friedrich II, to succeed Euler. Rather late in his life, in 1787, did Lagrange move to Paris, invited by Louis XVI to teach at *Ecole normale*. Appointed senator and made count by Napoleon, Lagrange became one of the first professors at *Ecole polytechnique*.

The first-order normality conditions for inequality-constrained problems had to wait until well into the XX century. These conditions were disclosed first by W. Karush in his M.S. thesis in the Department of Mathematics at the University of Chicago (Karush, 1939). Apparently, these results were never published in the archival literature, for which reason they remained unknown. Twelve years later, they were published in the *Proc. Second Berkeley Symposium* by H.W. Kuhn and A.W. Tucker (1951). The credit of these normality conditions has gone mostly to Kuhn and Tucker, but given their history, these conditions should be referred to as the *Karush-Kuhn-Tucker conditions*.

The problem at hand is formulated as

$$f(\mathbf{x}) \quad \rightarrow \quad \min_{\mathbf{x}} \tag{5.1a}$$

subject to

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}_m \tag{5.1b}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}_l \tag{5.1c}$$

where inequality (5.1b) is to be taken with a grain of salt: Arrays *not forming ordered sets*, this relation has no verbatim meaning. It is to be interpreted as shorthand for a set of m inequalities, namely,

$$\begin{aligned} g_1(\mathbf{x}) &\equiv g_1(x_1,\, x_2,\, \ldots,\, x_n) \le 0 \\ g_2(\mathbf{x}) &\equiv g_2(x_1,\, x_2,\, \ldots,\, x_n) \le 0 \\ &\vdots \\ g_m(\mathbf{x}) &\equiv g_m(x_1,\, x_2,\, \ldots,\, x_n) \le 0 \end{aligned}$$

To formulate the normality conditions, we proceed as before, namely, by defining a *Lagrangian* upon adjoining the equality and the inequality constraints to the objective function, namely,

$$F(\mathbf{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) \equiv f(\mathbf{x}) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T\mathbf{g}(\mathbf{x}) \quad \rightarrow \quad \min_{\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}} \tag{5.2}$$

While the normality conditions cannot be derived by simply making the gradient of the foregoing Lagrangian equal to zero, these conditions look very much like those associated with equality-constrained problems. Indeed, $\mathbf{x}_0$ is a stationary point if

$$\mathbf{h}(\mathbf{x}_0) = \mathbf{0}_m, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}_l \tag{5.3a}$$

$$\nabla f|_{\mathbf{x}=\mathbf{x}_0} + \mathbf{J}_0^T\boldsymbol{\lambda} + \mathbf{G}_0^T\boldsymbol{\mu} = \mathbf{0}_n \tag{5.3b}$$

$$\boldsymbol{\lambda} \ne \mathbf{0}_l, \quad \boldsymbol{\mu} \ge \mathbf{0}_m, \quad \boldsymbol{\mu}^T\mathbf{g}(\mathbf{x}) = 0 \tag{5.3c}$$

where

$$\mathbf{J}_0 \equiv \mathbf{J}(\mathbf{x}_0) \equiv \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}, \quad \mathbf{G}_0 \equiv \mathbf{G}(\mathbf{x}_0) \equiv \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0} \tag{5.3d}$$

i.e., $\mathbf{J}$ and $\mathbf{G}$ are, respectively, $l \times n$ and $m \times n$ matrices. Moreover,

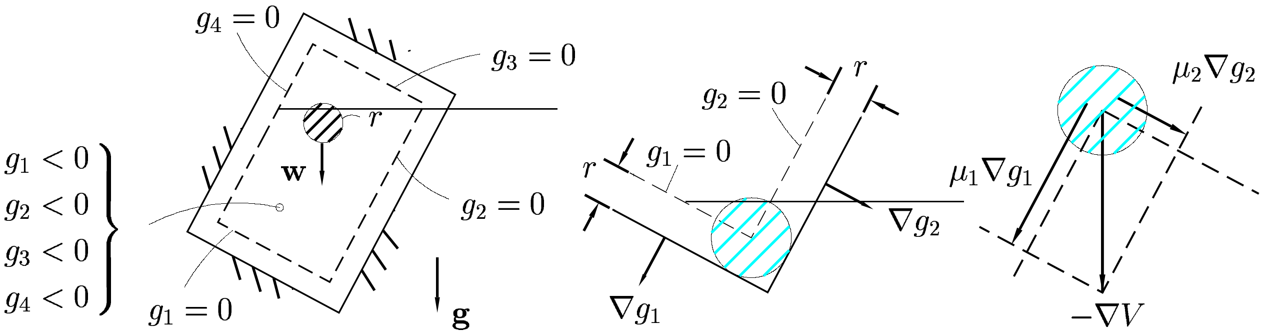
$$\mathbf{G} \equiv \begin{bmatrix} (\nabla g_1)^T \\ (\nabla g_2)^T \\ \vdots \\ (\nabla g_m)^T \end{bmatrix}, \quad \mathbf{J} \equiv \begin{bmatrix} (\nabla h_1)^T \\ (\nabla h_2)^T \\ \vdots \\ (\nabla h_l)^T \end{bmatrix} \tag{5.4}$$

Hence, the KKT condition (5.3b) can be expressed alternatively as

$$\begin{aligned} \nabla f|_{\mathbf{x}=\mathbf{x}_0} &+ \lambda_1 \nabla h_1|_{\mathbf{x}=\mathbf{x}_0} + \lambda_2 \nabla h_2|_{\mathbf{x}=\mathbf{x}_0} + \cdots + \lambda_1 \nabla h_1|_{\mathbf{x}=\mathbf{x}_0} \\ &+ \mu_1 \nabla g_1|_{\mathbf{x}=\mathbf{x}_0} + \mu_2 \nabla g_2|_{\mathbf{x}=\mathbf{x}_0} + \cdots + \mu_m \nabla g_m|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}_n \end{aligned} \tag{5.5}$$

Relations (5.3a–c) are known as the *Karush-Kuhn-Tucker (KKT) conditions.*

Remark: In the absence of inequality constraints, eq.(5.3a) reduces to the FONC of equality-constrained problems, eq.(3.23a).



(a) The general layout

(b) The ball at its equilibrium position

(c) A decomposition of the weight of the ball in this position

Figure 5.1: A heavy ball inside a box inclined with respect to the vertical

In the third row of the KKT, eqs.(5.3c), the third of these relations is sometimes referred to as the *transversality condition.* Upon expansion, this condition states

$$\mu_1 g_1(\mathbf{x}_0) + \mu_2 g_2(\mathbf{x}_0) + \cdots + \mu_m g_m(\mathbf{x}_0) = 0 \tag{5.6}$$

In light of the inequalities (5.1b) and the second relation of (5.3c), the ith transversality condition states that, at a stationary point, for $i = 1, 2, \ldots, m$, either $\mu_i = 0$ or $g_i(\mathbf{x}_0) = 0$. When, at a stationary point, the ith constraint of (5.1b) holds with the equality sign, this constraint is said to be *active.* Thus, *at a stationary point, the components of* $\boldsymbol{\mu}$ *associated with active constraints are positive, all other components vanishing.* As a consequence, if a of the m inequality constraints are active, we can partition vector $\mathbf{g}(\mathbf{x}_0)$, very likely after a reshuffling of the components of vector $\mathbf{g}$, in the form

$$\mathbf{g}(\mathbf{x}_0) = \begin{bmatrix} \mathbf{g}_a \\ \mathbf{g}_{m'} \end{bmatrix}$$

where $\mathbf{g}_a$ and $\mathbf{g}_{m'}$ are a- and $(m-a)$-dimensional vectors, respectively. Now, eq.(5.5) can be restated as

$$\begin{aligned} \nabla f|_{\mathbf{x}=\mathbf{x}_0} + \lambda_1 \nabla h_1|_{\mathbf{x}=\mathbf{x}_0} + \lambda_2 \nabla h_2|_{\mathbf{x}=\mathbf{x}_0} + \cdots + \lambda_1 \nabla h_1|_{\mathbf{x}=\mathbf{x}_0} \\ + \mu_1 \nabla g_1|_{\mathbf{x}=\mathbf{x}_0} + \mu_2 \nabla g_2|_{\mathbf{x}=\mathbf{x}_0} + \cdots + \mu_a \nabla g_a|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}_n, \; \mathbf{g}_{m'} = \mathbf{0}_{m'} \end{aligned} \tag{5.7}$$

The theoretical bases of the KKT conditions are to be found in the Farkas Lemma (Brousse, 1988). For the sake of conciseness, we do not elaborate on these issues.

To illustrate their validity, we give in Fig. 5.1 a mechanical interpretation of these conditions: A heavy ball of weight $\mathbf{w}$ is constrained to lie in a box, under the action of the gravity field $\mathbf{g}$, as depicted in Fig. 5.1a; the ball is shown in its equilibrium position in Fig. 5.1b; in Fig. 5.1c, the weight of the ball, equal to $-\nabla V$, where V is the potential energy of the ball, is decomposed into the two forces normal to the box walls. Notice that these two components push the walls, but cannot pull them, which is the reason why $\mu_i > 0$, for $i = 1, 2$.

Example 5.2.1

Consider the problem

$$f = \frac{1}{2}(x_1^2 + x_2^2) \quad \rightarrow \quad \min_{x_1, x_2}$$

subject to

$$x_1 + x_2 \geq 10$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

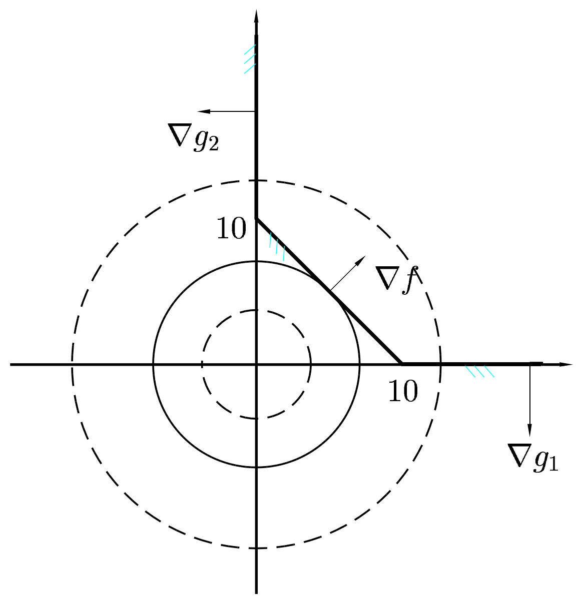


Figure 5.2: A quadratic objective function subject to linear inequality constraints

The objective function and the constraints are illustrated in Fig. 5.2.

For starters, we must express the inequality constraints in the standard form adopted at the outset, i.e., as $g_i(\mathbf{x}) \leq 0$*, whence,*

$$g_1 \equiv -x_1 - x_2 + 10 \leq 0, \quad g_2 \equiv -x_1 \leq 0, \quad g_3 \equiv -x_2 \leq 0$$

Apparently, the minimum is found at $\mathbf{x}_0 = [5,\ 5]^T$*. We evaluate then the items entering in the KKT conditions at this point* $\mathbf{x}_0$*:*

$$\nabla g_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla g_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \nabla g_3 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \Rightarrow \quad \nabla f|_{\mathbf{x}=\mathbf{x}_0} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Apparently, only the first constraint is active, and hence,

$$g_1(\mathbf{x}_0) = 0, \quad \mu_1 > 0, \quad \mu_2 = \mu_3 = 0$$

The KKT condition (5.7) thus reduces to

$$\nabla f|_{\mathbf{x}=\mathbf{x}_0} + \mu_1 \nabla g_1|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}_2$$

or

$$\mu_1 \nabla g_1|_{\mathbf{x}=\mathbf{x}_0} = - \nabla f|_{\mathbf{x}=\mathbf{x}_0}$$

which states that, at the SP $\mathbf{x}_0$ *given above, the two gradients,* $\nabla f|_{\mathbf{x}=\mathbf{x}_0}$ *and* $\nabla g_1|_{\mathbf{x}=\mathbf{x}_0}$*, are linearly-dependent. As a consequence, the above overdetermined system of two equations in one single unknown,* μ_1*, admits one solution that verifies the two equations. Upon solving this system, in fact, we obtain* $\mu_1 = 5 > 0$*, thereby verifying the second relation of conditions (5.3c).*

Example 5.2.2

$$f \equiv 8x_1^2 - 8x_1x_2 + 3x_2^2 \quad \rightarrow \quad \min_{\mathbf{x}}$$

subject to

$$x_1 \geq 3$$
$$x_2 \leq \frac{3}{2}$$

The objective function and the constraints of this example are depicted in Fig. 5.2.2

Again, we start by restating the inequalities in our standard form:

$$g_1 \equiv 3 - x_1 \leq 0, \; g_2 \equiv x_2 - \frac{3}{2}$$

Therefore,

$$\nabla g_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \; \nabla g_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

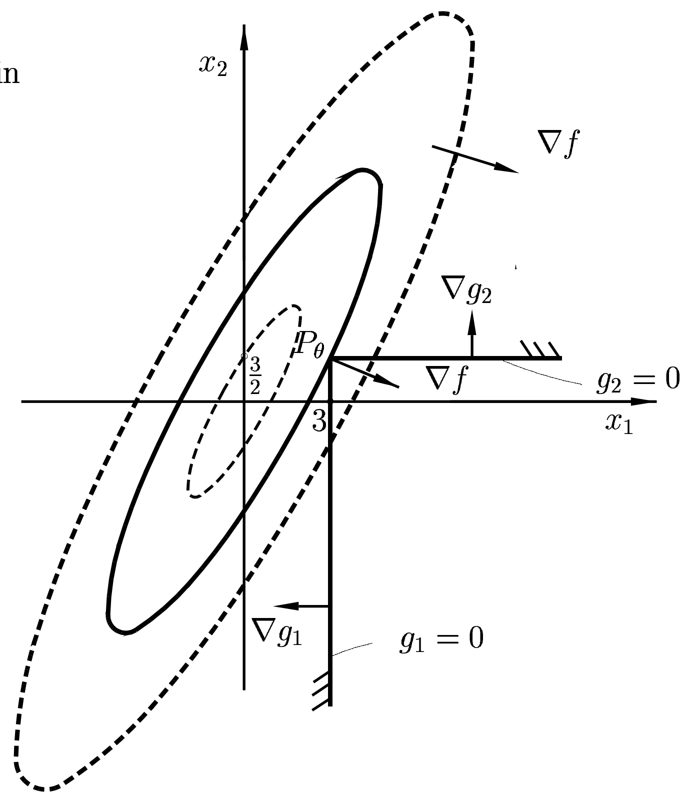


Figure 5.3: One more quadratic objective function subject to linear inequality constraints

condition (5.3b) thus leading to

$$\mu_1 \nabla g_1|_{\mathbf{x}=\mathbf{x}_0} + \mu_2 \nabla g_2|_{\mathbf{x}=\mathbf{x}_0} = -\nabla f|_{\mathbf{x}=\mathbf{x}_0}$$

where, apparently,

$$\mathbf{x}_0 = \begin{bmatrix} 3 \\ 3/2 \end{bmatrix}, \quad \nabla f = \begin{bmatrix} 16x_1 - 8x_2 \\ -8x_1 + 6x_2 \end{bmatrix}$$

Hence,

$$\nabla f|_{\mathbf{x}=\mathbf{x}_0} = \begin{bmatrix} 36 \\ 15 \end{bmatrix}$$

The above normality condition thus leading to

$$\mu_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -36 \\ 15 \end{bmatrix}$$

which, in this case, turns out to be a determined system of two equations in two unknowns, its solution being

$$\mu_1 = 36 > 0, \quad \mu_2 = 15 > 0$$

thereby verifying all KKT conditions.

Example 5.2.3 (A Linear Program)

A class of optimization problems finding a number of applications involves a linear objective function subject to linear equality and inequality constraints. This class is studied within the realm of linear programming. *These problems cannot be solved with the tools described so far, for we have focused on least-square problems, with an extension to more general objective functions and equality constraints. By the same token, linear programs arise seldom in mechanical design. To be true, a family of design problems in structural engineering, known as* limit design, *pertain to the design of structural elements, beams, columns and plates, for minimum weight, in such a way that all modes of* plastic failure *are avoided. Problems in limit design lead to linear programs.*

Linear programming is a first instance of application of the KKT conditions. We illustrate the concept with the problem below.

$$f \equiv 2x_1 - x_2 \quad \rightarrow \quad \min_{x_1, x_2}$$

subject to

$$g_1(\mathbf{x}) \equiv -x_1 \leq 0$$
$$g_2(\mathbf{x}) \equiv -x_2 \leq 0$$
$$g_3(\mathbf{x}) \equiv x_1 + x_2 - 1 \leq 0$$

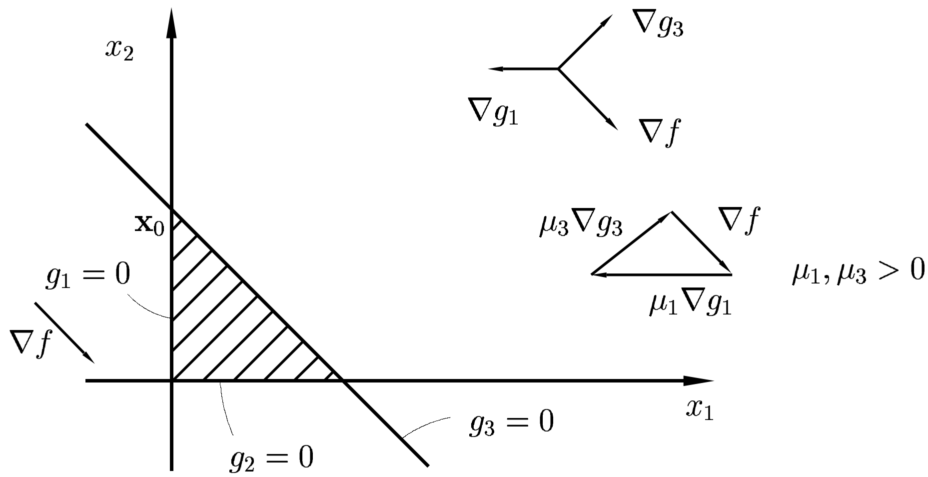


Figure 5.4: A linear program

The objective function and the constraints of this problem are illustrated in Fig. 5.4.

In this case,

$$\mathbf{G} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}, \quad \nabla f = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The KKT conditions lead to

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Apparently, g_1 and g_3 are active, and hence,

$$\mu_1,\, \mu_3 > 0, \quad \mu_2 = 0$$

thereby ending up with a system of two equations in two unknowns, μ_1 and μ_3. Upon solving this system, we obtain, successively,

$$\mu_3 = 1 > 0, \quad \mu_1 = 3 > 0$$

thereby verifying the KKT conditions.

5.3 Second-Order Normality Conditions

The simplest way of stating the *sufficient* conditions for a minimum, i.e., the second-order normality conditions, is by imposing the condition that, at a stationary point, any *feasible move* will produce an increment in the objective function, i.e., if we let $\Delta\mathbf{x}_F$ be a feasible move, then

$$\Delta f \equiv f(\mathbf{x}_0 + \Delta\mathbf{x}_F) - f(\mathbf{x}_0) > 0 \tag{5.8}$$

$$\Delta\mathbf{h} \equiv h(\mathbf{x}_0 + \Delta\mathbf{x}_F) - \mathbf{h}(\mathbf{x}_0) = \mathbf{0}_l \tag{5.9}$$

$$\Delta\mathbf{g} \equiv g(\mathbf{x}_0 + \Delta\mathbf{x}_F) - \mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}_m \tag{5.10}$$

Note that eq.(5.9) constrains $\Delta\mathbf{x}_F$ to lie in the nullspace of $\mathbf{J}$, i.e.,

$$\mathbf{J}\Delta\mathbf{x}_F = \mathbf{0}_l \tag{5.11}$$

and, if we resort to the isotropic orthogonal complement of $\mathbf{J}$ introduced in eq.(3.33), then a move that verifies the above condition is defined as

$$\Delta\mathbf{x}_F = \mathbf{L}\Delta\mathbf{u} \tag{5.12}$$

condition (5.8) then leading to

$$\mathbf{L}^T\nabla\nabla f\big|_{\mathbf{x}=\mathbf{x}_0}\mathbf{L} > \mathbf{O} \tag{5.13}$$

which states that the feasible Hessian must be positive-definite. However, in this case, contrary to that of Ch. 3, $\Delta\mathbf{u}$ is not free; it is constrained to obey conditions (5.10), i.e.,

$$\mathbf{G}^T\mathbf{L}\Delta\mathbf{u} \leq \mathbf{0}_m \tag{5.14}$$

If we let

$$\Delta\mathbf{v} \equiv \mathbf{L}\boldsymbol{\Delta}u \tag{5.15}$$

then, the above condition can be stated as

$$(\Delta v_1)\nabla g_1 + (\Delta v_2)\nabla g_2 + \cdots + (\Delta v_m)\nabla g_m \leq \mathbf{0}_m \tag{5.16}$$

Now, if we assume that only a of the m inequality constraints are active, the foregoing condition leads to

$$(\Delta v_1)\nabla g_1 + (\Delta v_2)\nabla g_2 + \cdots + (\Delta v_m)\nabla g_a \leq \mathbf{0}_a \tag{5.17}$$

where a reshuffling of the inequalities may have been needed.

5.4 Methods of Solution

Two classes of methods are available to solve inequality-constrained problems: a) *direct methods*, which handle the inequalities as such, and b) *indirect methods*, which transform the problem into one of the equality-constrained type. We will discuss only the latter.

Inequality-constrained problems can be solved using the approach introduced for either unconstrained or equality-constrained problems, upon converting the problem at hand into an unconstrained or, correspondingly, an equality constrained problem. This can be done by various methods: slack variables; interior or exterior penalty functions; etc. In this section, the methods of slack variables and interior penalty functions are outlined.

5.4.1 Slack Variables

Upon introducing the *slack variables* $s_1, s_2, \cdots, s_m$ into inequalities (5.1b), we convert these inequalities into equality constraints, namely,

$$\mathbf{h}(\mathbf{x},\mathbf{s}) \equiv \begin{bmatrix} g_1 + s_1^2 \\ g_2 + s_2^2 \\ \vdots \\ g_l + s_m^2 \end{bmatrix} = \mathbf{0}, \quad \mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{s} \equiv \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} \tag{5.18}$$

Notice that the slack variables being unknown, they have to be treated as additional design variables, the dimension of the design space being correspondingly increased. In consequence, the design vector is now of dimension $n+m$, i.e.,

$$\boldsymbol{\xi} \equiv \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} \tag{5.19}$$

Now, the gradient of the objective function with respect to the new design-variable vector takes the form

$$\nabla_{\boldsymbol{\xi}} f \equiv \begin{bmatrix} \nabla f \\ \nabla_{\mathbf{s}} f \end{bmatrix} \tag{5.20a}$$

where

$$\nabla f \equiv \frac{\partial f}{\partial \mathbf{x}}, \quad \nabla_{\mathbf{s}} f \equiv \frac{\partial f}{\partial \mathbf{s}} = \mathbf{0}_m \tag{5.20b}$$

the second relation following because the slack variable do not appear explicitly in the objective function.

Likewise, the Hessian w.r.t. the new design-variable vector $\boldsymbol{\xi}$ takes the form

$$\nabla\nabla_{\boldsymbol{\xi}} f = \begin{bmatrix} \nabla\nabla f & \partial^2 f/\partial\mathbf{x}\partial\mathbf{s} \\ (\partial^2 f/\partial\mathbf{x}\partial\mathbf{s})^T & \nabla\nabla_{\mathbf{s}} f \end{bmatrix} \tag{5.21a}$$

with the notation

$$\nabla\nabla_{\mathbf{s}} \equiv \frac{\partial^2 f}{\partial \mathbf{s}^2} \tag{5.21b}$$

However, since $\nabla_{\mathbf{s}} = \mathbf{0}_m$, the above Hessian expression reduces to

$$\nabla\nabla_{\boldsymbol{\xi}} f = \begin{bmatrix} \nabla\nabla f & \mathbf{O}_{nm} \\ \mathbf{O}_{nm}^T & \mathbf{O}_m \end{bmatrix} \tag{5.22}$$

That is, the Hessian of the objective function w.r.t. the new design-variable vector $\boldsymbol{\xi}$ is singular. In case $\nabla\nabla f$ is positive-definite, $\nabla\nabla_{\boldsymbol{\xi}} f$ is positive-semidefinite. Hence, in applying the method of slack variables to solve inequality-constrained problems, Hessian stabilization will always be needed.

Example 5.4.1 (Minimization of the Design Error of a Four-Bar Linkage with an Input Crank)
Determine the link-lengths of the four-bar linkage shown in Fig. 5.5, that will produce the set of input-output pairs $\{\psi_i, \phi_i\}_1^q$ shown in Table 5.1, where ψ and ϕ denote the input and output angles.

Table 5.1: The input-output pairs of $\{\psi_i, \phi_i\}_1^{10}$

i	1	2	3	4	5
ψ_i	123.8668°	130.5335°	137.2001°	143.8668°	150.5335°
ϕ_i	91.7157°	91.9935°	92.8268°	94.2157○	96.1601°
i	6	7	8	9	10
ψ_i	157.2001°	163.8668°	170.5335°	177.2001°	183.8668°
ϕ_i	98.6601°	101.7157°	105.3268°	109.4935°	114.2157°

The link-lengths are obtained via the Freudenstein parameters *k_1, k_2 and k_3, defined as*

$$k_1 = \frac{a_1^2 + a_2^2 - a_3^2 + a_4^2}{2a_2a_4}, \quad k_2 = \frac{a_1}{a_2}, \quad k_3 = \frac{a_1}{a_4} \tag{5.23a}$$

with the inverse relations

$$a_2 = \frac{a_1}{k_1}, \quad a_3 = \frac{\sqrt{k_2^2 + k_3^2 + k_2^2k_3^2 - 2k_1k_2k_3}}{|k_2k_3|}, \quad a_4 = \frac{a_1}{k_3} \tag{5.23b}$$

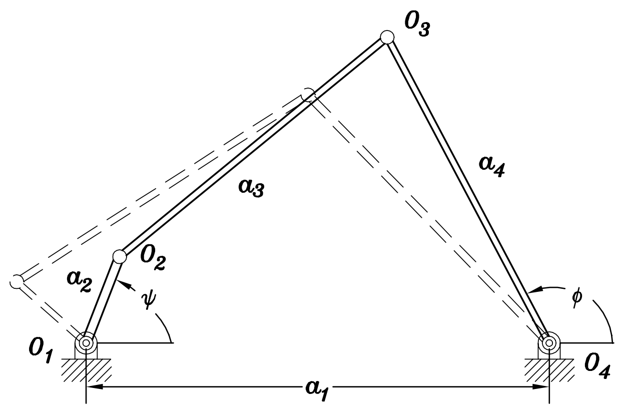


Figure 5.5: A four-bar linkage

for a given value of a_1. *The synthesis equations for the planar four-bar linkage can be written in the form (Liu and Angeles, 1993):*

$$\mathbf{Sk} = \mathbf{b} \tag{5.24}$$

where $\mathbf{S}$ *is the* synthesis matrix, *and* $\mathbf{k}$ *is the vector of linkage parameters. Moreover,* $\mathbf{S}$, $\mathbf{k}$ *and* $\mathbf{b}$ *are defined as*

$$\mathbf{S} = \begin{bmatrix} 1 & \cos\phi_1 & -\cos\psi_1 \\ 1 & \cos\phi_2 & -\cos\psi_2 \\ \vdots & \vdots & \vdots \\ 1 & \cos\phi_q & -\cos\psi_q \end{bmatrix} \quad \mathbf{k} = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} \cos(\psi_1 - \phi_1) \\ \cos(\psi_2 - \phi_2) \\ \vdots \\ \cos(\psi_q - \phi_q) \end{bmatrix} \tag{5.25}$$

The design error *is defined as*

$$\mathbf{d} \equiv \mathbf{b} - \mathbf{Sk} \tag{5.26}$$

the purpose of the optimization exercise being to minimize the Euclidean norm of the design error, while ensuring that its input link is a crank. The conditions for full mobility *of the input crank are (Liu and Angeles, 1992)*

$$g_1(\mathbf{x}) = (k_1 + k_3)^2 - (1 + k_2)^2 < 0$$
$$g_2(\mathbf{x}) = (k_1 - k_3)^2 - (1 - k_2)^2 < 0$$

By introducing two slack-variables s_1 *and* s_2, *the inequality constraints are converted into equality constraints, i.e.,*

$$h_1(\mathbf{x}) = (k_1 + k_3)^2 - (1 + k_2)^2 + s_1^2 = 0 \tag{5.27a}$$
$$h_2(\mathbf{x}) = (k_1 - k_3)^2 - (1 - k_2)^2 + s_2^2 = 0 \tag{5.27b}$$

The design vector $\boldsymbol{\xi}$ *thus becomes* $\boldsymbol{\xi} = [k_1 \; k_2 \; k_3 \; s_1 \; s_2]^T$. *From the initial guess* $\boldsymbol{\xi}_0 = [0.28 \; 0.74 \; 0.12 \; 1.69 \; 1.2]^T$, *the solution was found to be* $\boldsymbol{\xi}_{\text{opt}} = [0.3248 \;\; 0.5875 \; -0.009725 \; 1.556 \; 0.2415]^T$, *and the corresponding link lengths are* $a_1 \equiv 1$, $a_2 = 1.702$, $a_3 = 103.4$ *and* $a_4 = 102.8$. *The Euclidean norm of the minimum design error is* 5×10^{-2}.

The problem with this design is that it leads to a quite disproportionate linkage: two of its links have lengths two orders of magnitude bigger than those of the other two!

5.4.2 Interior Penalty Functions

The idea behind penalty-function methods is to approach the optimum solution *asymptotically*, by extrapolation of a sequence of optimum solutions to unconstrained problems. There are two possibilities: the solution is approached either *within* the feasible region or from *without*, the penalty function being correspondingly referred to as *interior* or *exterior*. It is noteworthy that exterior penalty-function methods are applicable *only* to problems whereby the optimum finds itself at the boundary of the feasible region, but misses interior optima. Hence, we focus here on interior penalty functions.

Given an objective function $f(\mathbf{x})$ subject to inequality constraints, as defined in eq.(5.1b), a sequence of interior penalty functions $\{\phi_k\}_1^\nu$ is constructed as

$$\phi_k(\mathbf{x}; r_k) \equiv f(\mathbf{x}) + r_k \sum_{i=1}^{m} \frac{1}{g_i(\mathbf{x})} \quad k = 1, 2, \cdots, \nu \tag{5.28}$$

where all r_k's are positive and observe a decreasing order, i.e.,

$$r_1 > r_2 > r_3 \cdots > r_\nu \tag{5.29}$$

Now, a sequence of unconstrained optimization problems is defined:

$$\phi_k(\mathbf{x}; r_k) \equiv f(\mathbf{x}) + r_k \sum_{i=1}^{m} \frac{1}{g_i(\mathbf{x})} \quad \rightarrow \quad \min_{\mathbf{x}}, \quad k = 1, 2, \cdots, \nu \tag{5.30}$$

Let $\mathbf{x}_o^1$, $\mathbf{x}_o^2$, ..., $\mathbf{x}_o^\nu$ be the sequence of corresponding unconstrained optima. Next, these optima are interpolated to a vector function $\mathbf{x}_o(r)$:

$$\mathbf{x}_o(r) \equiv \mathbf{c}_0 + \sum_{1}^{\nu-1} \mathbf{c}_k r^{k/2} \tag{5.31}$$

thereby obtaining a system of νn equations in νn unknowns, the n components of the ν unknown vector coefficients $\{\,\mathbf{c}_k\,\}_0^{\{\nu-1\}}$. Note that the foregoing equations are all linear in the unknowns, and hence, they can be solved for the unknowns using Gaussian elimination, as described below.

First, eq.(5.31) is written for $r = r_i$, with $i = 1,\, 2,\, \ldots,\, \nu$:

$$\mathbf{x}_o(r_i) \equiv \mathbf{c}_0 + \mathbf{c}_1 r_i^{1/2} + \mathbf{c}_2 r_i^{2/2} + \cdots + \mathbf{c}_{\nu-1} r_i^{(\nu-1)/2} \tag{5.32}$$

or

$$\mathbf{x}_o(r_i) \equiv [\,\mathbf{c}_0 \quad \mathbf{c}_1 \quad \cdots \quad \mathbf{c}_{\nu-1}\,] \begin{bmatrix} 1 \\ r_i^{1/2} \\ \vdots \\ r_i^{(\nu-1)/2} \end{bmatrix}, \qquad i = 1,\, 2,\, \ldots,\, \nu$$

In the next step, we regroup all ν vector equations above to produce a matrix equation. To this end, we define the matrices

$$\mathbf{R} \equiv \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1^{1/2} & r_2^{1/2} & \cdots & r_\nu^{1/2} \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{(\nu-1)/2} & r_2^{(\nu-1)/2} & \cdots & r_\nu^{(\nu-1)/2} \end{bmatrix} \tag{5.33a}$$

$$\mathbf{X}_o \equiv [\,\mathbf{x}_o(r_1) \quad \mathbf{x}_o(r_2) \quad \mathbf{x}_o(r_\nu)\,] \tag{5.33b}$$

$$\mathbf{C} \equiv [\,\mathbf{c}_0 \quad \mathbf{c}_1 \quad \cdots \quad \mathbf{c}_{\nu-1}\,] \tag{5.33c}$$

Thus, the ν vector equations (5.32) become, in matrix form,

$$\mathbf{C}\mathbf{R} = \mathbf{X}_o \tag{5.34a}$$

whence,

$$\mathbf{C} = \mathbf{X}_o\mathbf{R}^{-1} \tag{5.34b}$$

or, if eq.(5.34a) is written in the usual form, with the unknown matrix $\mathbf{C}$ to the left of its matrix coefficient, the foregoing equation should first be transposed, the result then being

$$\mathbf{C}^T = \mathbf{R}^{-T}\mathbf{X}_o^T \tag{5.34c}$$

with exponent $-T$ indicating the inverse of the transpose or, equivalently, the transpose of the inverse. Once the ν vector coefficients sought are available, the optimum of the inequality-constrained problem, $\mathbf{x}_{\text{opt}}$, is calculated as

$$\mathbf{x}_{\text{opt}} = \lim_{r\to 0} \mathbf{x}_{\text{opt}}(r)$$

i.e.,

$$\mathbf{x}_{\text{opt}} = \mathbf{c}_0 \tag{5.35}$$

In computing the above value, note that $\mathbf{c}_0$ is the first column of the unknown matrix $\mathbf{C}$ or, equivalently, the first row of its transpose. In either case, it is not possible to obtain $\mathbf{c}_0$ as the solution of one single vector equation. A matrix equation must be solved in order to obtain $\mathbf{c}_0$.

Example 5.4.2 (A Two-dimensional Optimization Problem Subject to Inequality Constraints) *Consider an optimization problem with an objective function defined as*

$$f = x^2 + 2y^2 \quad \rightarrow \quad \min_{x,y} \tag{5.36}$$

subject to inequality constraints

$$g_1 \equiv -x \leq 0 \tag{5.37a}$$
$$g_2 \equiv -y \leq 0 \tag{5.37b}$$
$$g_3 \equiv 1 - x - y \leq 0 \tag{5.37c}$$

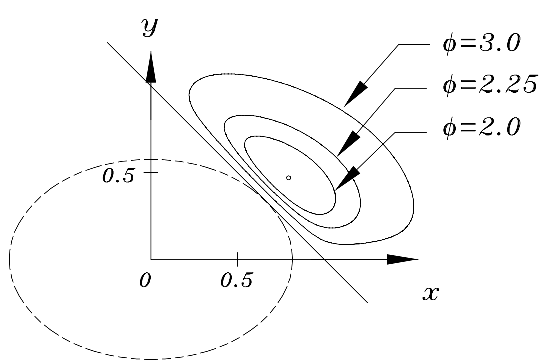


Figure 5.6: Isocontours of the penalty function with $r_1 = 0.1$

From a sketch of the constraints and the contours of the objective function in the x-y plane, it should be apparent that the minimum of f is attained at a point where the gradient ∇f is parallel to the normal to the line $g_3(x, y) = 0$. The optimum values of x and y are, then

$$x_{\text{opt}} = \frac{2}{3}, \quad y_{\text{opt}} = \frac{1}{3}$$

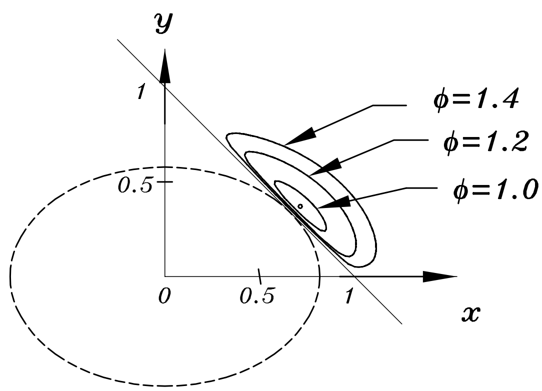


Figure 5.7: Isocontours of the penalty function with $r_2 = 0.01$

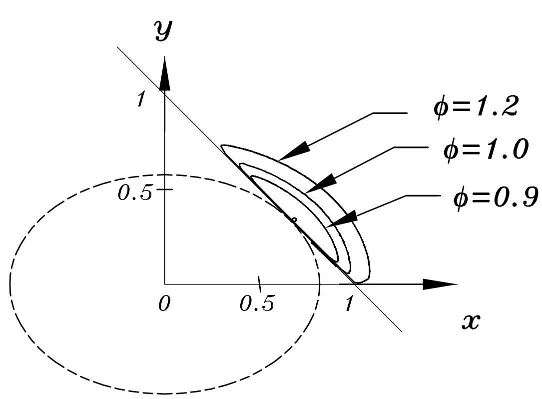


Figure 5.8: Isocontours of the penalty function with $r_3 = 0.001$

We demonstrate below the application of penalty functions to obtain the foregoing optimum. We have

$$\phi_k \equiv x^2 + 2y^2 + r_k \left(\frac{1}{x} + \frac{1}{y} + \frac{1}{x+y-1} \right), \quad k = 1, \ldots, 3 \quad \rightarrow \quad \min_{x,y}$$

subject to no constraints, for

$$r_1 = 0.1, \quad r_2 = 0.01, \quad r_3 = 0.001$$

The penalty-function isocontours for different r_k *values are shown in Figs. 5.6–5.8. In those figures, the isocontour of the objective function* f *that corresponds to the constrained minimum is indicated with a dashed curve.*

The optima $\mathbf{x}_o(r_k) \equiv [\, x_o(r_k),\, y_o(r_k)\,]^T$ *for the three given values of* r_k *were found by the ODA as*

$$\mathbf{x}_o(r_1) = \begin{bmatrix} 0.7941 \\ 0.4704 \end{bmatrix}, \quad \mathbf{x}_o(r_2) = \begin{bmatrix} 0.7140 \\ 0.3703 \end{bmatrix}, \quad \mathbf{x}_o(r_3) = \begin{bmatrix} 0.6836 \\ 0.3434 \end{bmatrix} \tag{5.38}$$

We now fit the values of $\{\mathbf{x}_o(r_k)\}_1^3$ *to the function*

$$\mathbf{x}_o(r) = \mathbf{c}_0 + \mathbf{c}_1 r^{1/2} + \mathbf{c}_2 r$$

We thus have

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 1 \\ 0.3163 & 0.1000 & 0.03163 \\ 0.1000 & 0.0100 & 0.00100 \end{bmatrix}, \quad \mathbf{X}_o = \begin{bmatrix} 0.7941 & 0.7140 & 0.6836 \\ 0.4704 & 0.3703 & 0.3434 \end{bmatrix} \tag{5.39}$$

The coefficient matrix $\mathbf{C}$ *is thus found to be*

$$\mathbf{C} = \mathbf{X}_o \mathbf{R}^{-1} = \begin{bmatrix} 0.6687 & 0.4790 & -0.2605 \\ 0.3317 & 0.3612 & 0.2443 \end{bmatrix} \tag{5.40}$$

Therefore,

$$\mathbf{x}_{\text{opt}} = \mathbf{c}_0 = \begin{bmatrix} 0.6687 \\ 0.3317 \end{bmatrix} \tag{5.41}$$

which yields the optimum with two significant digits of accuracy.

References

Angeles, J., 1982, *Spatial Kinematic Chains*, Springer-Verlag, New York.

Angeles, J., Anderson, K. and Gosselin, C., 1990, "Constrained design optimization using orthogonal decomposition", *ASME J. Mechanisms, Transmissions, and Auto. in Design*, Vol. 112, No. 2, pp. 255-256.

Bertsekas, D. P., 1995, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts.

Betts, J. T., 1980, "A compact algorithm for computing the stationary point of a quadratic function subject to linear constraints", *ACM Trans. on Math. Software*, Vol. 63, pp. 391-397.

Broyden, C. G., 1970, "The convergence of a class of double-rank minimization algorithm", *J. Inst. Maths. Applics.*, Vol. 6, pp. 76-90.

Brousse, P., *Optimization in Mechanics: Problems and Methods*, North-Holland, Amsterdam.

Fox, R. L. and Gupta, K. C., 1973, "Optimization technology as applied to mechanism design", *ASME Journal of Engineering for Industry*, pp. 657-661.

Freudenstein, F., 1955, "Approximate synthesis of four-bar linkages", *Trans. ASME*, Vol. 77, pp. 853-861.

Goldfarb, D., 1970, "A family of variable metric updates derived by variational means", *Mathematics of Computing*, Vol. 24, pp. 23-26.

Golub, G. H. and Van Loan, C. F., 1983,*Matrix Computations*, The Johns Hopkins University Press, Baltimore.

Hartenberg R. S. and Denavit, J., 1964, *Kinematic Synthesis of Linkages*, McGraw-Hill Book Co., New York.

Kahaner D., Moler C. and Nash S., 1989, *Numerical methods and software*, Prentice Hall, Englewood Cliffs, NJ.

Karush, W., 1939, *Minima of Functions of Several Variables with Inequalities as Side Conditions*, M.S. Thesis, Department of Mathematics, University of Chicago, Chicago.

Kuhn, H.W. and Tucker, A.W., 1951, "Nonlinear Programming," in J. Neyman (editor), *Proc. Second Berkeley Simposium*, University of California Press, Berkeley, pp. 481–492.

Lalee, M., Nocedal, J. and Plantenga, 1998, "On the implementation of an algorithm for large-scale equality constrained optimization", *SIAM J. Optim.*, Vol. 8, pp. 682-706.

Liu, Z. and Angeles, J., 1992, "Least-square optimization of planar and spherical four-bar function generator under mobility constraints", *ASME, J. of Mech. Des.*, Vol. 114, pp. 569-573.

Mangasarian, O. L., 1972, "Techniques of optimization", *Trans. ASME Journal of Engineering for Industry*, Vol. 93, pp. 365-371.

Murray, W., 1997, "Sequential quadratic programming methods for large problems", *Comput. Optim. Appl.*, Vol. 7, pp. 127-142.

Powell, M. J. D., 1969, "A method for Nonlinear Constraints in Minimizing Problems", in Fletcher, R. (Ed.), *Optmization*, Academic Press., N. Y., pp. 283-298.

Rao, S. S., 1996, *Engineering Optimization*, 3rd. Edition, John Wiley and Sons, Inc., New York.

Rosenbrock, H. H., 1960, "An automatic method for finding the greatest or the least value of a function", *Computer Journal*, Vol. 3. No. 3, pp. 175-184.

Schittkowski, K., 1985, "A unified outline of nonlinear programming algorithms", *ASME Journal of Mechanisms, Transmissions, and Automation in Design*, Vol. 107, pp. 449-453.

Seireg, A., 1972, “A survey of optimization of mechanical design”, *ASME Journal Engineering for Industry*, Vol. 94, pp. 495-499.

Shanno, D. F., 1970, “Conditioning of quasi-Newton methods for function minimization”, *Mathematics of Computing*, Vol. 24, pp. 647-656.

Stewart, G. W., 1973, *Introduction to Matrix Computations, Academic Press*, New York.

Strang, G., 1988, *Linear Algebra*, Third Edition, Harcourt Brace Jovanovich College Publishers, Fort Worth.

Teng, C.P. and Angeles, J., 2001, “A sequential-quadratic-programming algorithm using orthogonal decomposition with Gerschgorin stabilization,” *ASME J. Mechanical Design*, Vol. 123, pp. 501–509.
Varga, R. S., 1962, *Matrix Iterative Analysis*, Prentice-Hall Inc., Englewood Cliffs, N. J..

Zhou, J. and Mayne, R. W., 1985, “Monotonicity analysis and recursive quadratic programming in constrained optimization”, *ASME Journal of Mechanisms, Transmissions, and Automation in Design*, Vol. 107, pp. 459-462.