1. Parameter Optimization Problems:

1.3. Inequality Constrained Optimization

©Erik I. Verriest*

November 19, 2008

1 Kuhn-Tucker Problem

We will now develop the necessary conditions for $y^* \in \mathbf{R}^n$ to be a local minimum for the function $L(y)$, with the p-dimensional constraint $f(y) \leq 0$. We denote by $x \leq 0$, for a vector x, the scalar conditions $x_i \leq 0$, $i = 1, \ldots p$.

If for some of the constraints $f_1(y^*) < 0$, we say that the corresponding constraint is *inactive*, and it may be eliminated from the set. If on the other hand $f_i(y) = 0$, then this constraint is said to be *active*. Clearly if at y^* all constraints are inactive, the necessary condition for optimality is that $L_y = 0$ as before.

Consider the more interesting case where $0 < m \leq p$ of the p given constraints are active. Consequently, we'll only look at these m constraints, and with some abuse of notation, we'll denote this m-vector of constraints again by $f(y)$. At this point m may be larger or smaller than n. We will also assume that the rank of f at y^* is r. This means that at the point $y*$, the $m \times n$ gradient matrix satisfies

$$\text{rank } f_y(y^*) = r \leq \min(m, n)$$

Equivalently, this expresses that the gradients of the effective constraints span an r-dimensional subspace of $\mathbf{R}^n$.
In what follows we shall also make use of the class of *selection matrices*. An $r - m$ selection matrix Γ is an $r \times m$ matrix whose rows are arbitrarily chosen from the $m \times m$ identity matrix. There always exists a selection matrix Γ such that rank $\Gamma f_y \, f_y' \Gamma' = r$.
If $r < n$, denote by Γ^c the complementary selection matrix, so that the rows of f are a permutation of the rows of $\begin{bmatrix} \Gamma f_y \\ \Gamma^c f_y \end{bmatrix}$.

*School of ECE, Georgia Tech, Atlanta, GA, USA. Presently with SCD, ESAT, KULeuven.

First we will derive the set of *admissible* perturbations dy. A suitable basis for the tangent space at y^*, i.e., the vector space attached at y^* (or with origin at y^*), are the r selected gradients (the rows of Γf_y) augmented (if $r < n$) with $n - r$ independent row vectors η in the orthogonal complement of the rowspace of Γf_y. Denoting this space by $f_y^\perp$, we also have that $f_y^\perp = [\Gamma f_y]^\perp$. Recall that vectors in the tangent space are row vectors. These rows, η, satisfy $\Gamma f_y \eta' = 0$. Changing back to columns (more precisely: the dual space), any arbitrary (column)-vector dy of dimension n can then be expressed as

$$dy = (\Gamma f_y)'(\Gamma f_y f_y' \Gamma')^{-1}\epsilon + \eta' \quad ; \quad \Gamma f_y \eta' = 0 \tag{1}$$

where $\epsilon \in \mathbf{R}^r$. Note that $(\Gamma f_y f_y' \Gamma')^{-1}\epsilon$ selects a particular linear combination of the columns of $(\Gamma f_y)'$. The vector η may also be expressed as

$$\eta = f_y^{\perp'} \nu, \tag{2}$$

with $\nu \in \mathbf{R}^{n-r}$. If $r = n$, then $f_y^\perp = 0$. This may seem like a strange parametrization, but the reason will become clear shortly. Admissibility of the perturbation, requires that the corresponding changes in f satisfy the constraint, or $df \leq 0$. Thus

$$df = f_y dy = f_y(\Gamma f_y)'(\Gamma f_y f_y' \Gamma')^{-1}\epsilon + f_y f_y^{\perp'} \nu \tag{3}$$

from which:

$$\begin{aligned} \Gamma df &= \epsilon, \quad (4)\\ \Gamma^c df &= 0, \quad (5) \end{aligned}$$

since $f_y f_y^{\perp'} = 0$ by definition. The condition for *admissibility* is thus $\epsilon \leq 0$.
What is now the effect on dL of such an admissible perturbation? We express the gradient L_y also in terms of the basis in the tangent space. Let thus for some row vectors λ and μ

$$L_y = -\lambda \Gamma f_y + \mu f_y^\perp \tag{6}$$

be the expansion of the gradient L_y in the chosen basis for the tangent space. Then the change in L due to the *admissible* perturbation dy is

$$\begin{aligned} dL &= L_y dy = [-\lambda \Gamma f_y + \mu f_y^\perp][(\Gamma f_y)'(\Gamma f_y f_y' \Gamma')^{-1}\epsilon + \eta'] \\ &= -\lambda\epsilon + \mu f_y^\perp \eta' \\ &= -\lambda\epsilon + \mu f_y^\perp f_y^{\perp'} \nu \quad (7) \end{aligned}$$

The two other terms vanish because $f_y f_y^\perp = 0$. As ν is arbitrary, the second term can have either sign unless $\mu = 0$. If y^* is a stationary solution to the problem IC, then at y^*, dL is necessarily nonnegative for all *admissible* excursions. The increment dL is positive for all admissible excursions dy if $\lambda \geq 0$ and $\mu = 0$. The second condition is equivalent to $L_y \in \text{rowspan} f_y$, while the first further restricts L_y to be in the *cone* spanned by the rows of $-f_y$ (the negative gradients of the constraints). This means that $L_y = -\lambda f_y$, where λ

cannot have negative components. This proves the following.

Kuhn-Tucker Theorem

The necessary conditions for y^* to minimize $L(y)$, subject to the m constraints, $f_y \leq 0$, are that

$$\begin{aligned} L_y(y^*) + \lambda f_y(y^*) &= 0 \\ \lambda f(y^*) &= 0 \\ \lambda &\geq 0 \text{ with } \lambda_i = 0 \text{ if } f_i(y^*) < 0. \end{aligned}$$