Chapter 9

The Method of Multipliers

9.1 Introduction

In our study of exterior penalty functions in the previous section we found that there was a compromise between differentiability and exactness. That is given a penalty function

$$P_\alpha(x) = f_0(x) + \alpha\beta(x)$$

either the penalty term is differentiable in which case the penalty parameter α must tend to $+\infty$ or $\beta(x)$ is nondifferentiable in which case α need not tend to $+\infty$. In this section we consider a modification to the quadratic differentiable penalty term $\hat{\beta}_2$ which avoids the need to send α to $+\infty$.

Recall that in each step of the exterior penalty method applied to

$$P_\alpha(x) := f_0(x) + \alpha\left[\sum_{i=1}^{s}(f_i(x)_+)^2 + \sum_{i=s+1}^{m}(f_i(x))^2\right]$$

one solves the unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} P_\alpha(x).$$

The solution x_α satisfies

$$(9.1.1) \qquad 0 = \nabla P_\alpha(x_\alpha) = \nabla f_0(x_\alpha) + \sum_{i=1}^{s}\nabla f_i(x_\alpha)(\alpha f_i(x_\alpha)_+) + \sum_{i=s+1}^{m}\nabla f(x_\alpha)(\alpha f_i(x)).$$

Setting

$$(u_\alpha)_i := \begin{cases} \alpha f_i(x_\alpha)_+ & \text{if } i = 1,\ldots,s \\ \alpha f_i(x_\alpha) & \text{if } i = s+1,\ldots,m \end{cases}$$

we can write (9.1.1) as

$$0 = \nabla_x L(x_\alpha, u_\alpha)$$

where $L(x,u) = f_0(x) + u^T f(x)$ is the Lagrangian for the problem $\mathcal{P}$.

$$\begin{array}{lll} \min & f_0(x) & \\ \text{subject to} & f_0(x) \le 0 & i = 1, \ldots, s \\ & f_0(x) = 0 & i = s+1, \ldots, m \end{array}$$

Consequently, if $x_\alpha \to \overline{x}$ a local solution to $\mathcal{P}$, then every cluster point of $\{u_\alpha\}$ is a Kuhn-Tucker multiplier for $\overline{x}$. This indicates that we should think of the vectors u_α as multiplier approximates that are to be updated at each iteration. By doing so, we avoid the need to send the penalty parameter α to $+\infty$. The strategy is as follows: given $\alpha > 0$ and an estimate of the multipliers $u \in \mathbb{R}^m$, let $x_{\alpha,u}$ be the solution to the equation

$$\begin{array}{rcl} 0 & = & \nabla f_0(x) + \sum_{i=1}^{s} \nabla f_i(x)(\alpha f_i(x) + u_i)_+ \\ & & + \sum_{i=s+1}^{m} \nabla f_i(x)(\alpha f_i(x) + u_i) \ . \end{array}$$

Then update the multiplier estimates u_i via the equations

$$u_i = (\alpha f_i(x_{\alpha,u}) + u_i)_+ \quad \text{for } i = 1, \ldots, s$$

and

$$u_i = (\alpha f_i(x_{\alpha,u}) + u_i) \quad \text{for } i = s+1, \ldots, m.$$

This procedure describes the basic structure of an algorithm known as *the method of multipliers*. Before we provide a precise description of this algorithm let us first examine expression (9.1) more carefully.

9.2 The Augmented Lagrangian

Observe that expression (9.1) is a first order optimality condition for some function. In order to recover this function we can integrate the right hand side of (9.1) in the variable x. By adding in the appropriate constant term this integration yields the function

$$\begin{array}{rcll} L(x,u,\alpha) & := & f_0(x) & + \frac{1}{2\alpha}[\text{dist}_2^2[\alpha f(x) + u|K] - \|u\|_2^2] \\ & := & f_0(x) & + \frac{1}{2\alpha}\sum_{i=1}^{s}((\alpha f_i(x) + u_i)_+)^2 - u_i^2 \\ & & & + \frac{\alpha}{2}\sum_{i=s+1}^{m} f_i(x)(f_i(x) + u_i). \end{array}$$

where $K := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$. The function $L(x,u,\alpha)$ is called the augmented Lagrangian for $\mathcal{P}$. The name is derived from the fact that if $s = 0$, that is there are only equality constraints, then $L(x,u,\alpha)$ takes the form

$$L(x,u,\alpha) = L(x,u) + \frac{\alpha}{2}\|f(x)\|_2^2$$

where $L(x,u) = f_0(x) + u^T f(x)$ is the usual Lagrangian. Thus $L(x,u,\alpha)$ can be thought of as arising from the usual Lagrangian after one has incorporated a vehicle for penalizing constraint violation. The augmented Lagrangian possesses the following remarkable property.

THEOREM 9.2.1 *Let $\alpha > 0$, f_i, $i = 0, \dots, m$ differentiable at $x \in \mathbb{R}^n$. Then*

$$0 = \nabla_{x,u} L(x, u, \alpha)$$

if and only if (x, u) is a Kuhn-Tucker pair for $\mathcal{P}$.

PROOF: Note that $0 = \nabla_{x,u} L(x, u, \alpha)$ if and only if

$$\begin{aligned} 0 &= \nabla f_0(x) + \sum_{i=1}^{s} (\alpha f_i(x) + u_i)_+ \nabla f_i(x) + \sum_{i=s+1}^{m} (\alpha f_i(x) + u_i) \nabla f_i(x) \\ u_i &= (\alpha f_i(x) + u_i)_+ \quad i = 1, \dots, s \\ 0 &= f_i(x) \quad i = s+1, \dots, m\,. \end{aligned}$$

Hence the result will be established once we have shown that

$$[(a - b)_+ - a = 0 \Longleftrightarrow a \geq 0, b \geq 0, ab = 0].$$

Case 1: $a - b \geq 0$
If $a - b \geq 0$, then $(a - b)_+ = a - b$ so that $b = 0$. Consequently, $a \geq 0$, $b \geq 0$, $ab = 0$.
Case 2: $a - b \leq 0$
If $(a - b) \leq 0$, then $(a - b)_+ = 0$ so that $a = 0$. Consequently, $a \geq 0$, $b \geq 0$ and $ab = 0$.
The converse is trivial. ■

Thus it would seem that we need only find the roots of the equation $0 = \nabla_{x,u} L(x, u, \alpha)$ in order to locate Kuhn-Tucker points for the problem $\mathcal{P}$. This is precisely what the method of multipliers attempts to do.

In order to investigate the rate of convergence for these methods we require the nonsingularity of the hessian

$$\nabla^2_{x,u} L(x, u, \alpha).$$

Unfortunately, $\nabla^2_{x,u} L(x, u, \alpha)$ does not always exists since $(\alpha f_i(x) + u_i)_+$ is not everywhere differentiable. A sufficient condition under which $\nabla^2_{x,u} L(x, u, \alpha)$ does exist near a Kuhn-Tucker point $(\overline{x}, \overline{u})$ for $\mathcal{P}$ is *strict complementary slackness.*

DEFINITION 9.2.1 *Let $(\overline{x}, \overline{u})$ be a Kuhn-Tucker pair for $\mathcal{P}$. We say that the strict complementary slackness condition (SCSC) is satisfied at $(\overline{x}, \overline{u})$ if $\overline{u}_i > 0$ whenever $f_i(\overline{x}) < 0$ $i = 1, \dots, s$.*

Observe that if the SCSC is satisfied at the K-T pair $(\overline{x}, \overline{u})$ then

$$(\alpha f_i(x) + u_i)_+ = 0$$

for all (x, u) near $(\overline{x}, \overline{u})$ for each $i \notin A(\overline{x}) = \{i : f_i(\bar{x}) = 0\}$, and

$$(\alpha f_i(x) + u_i)_+ = (\alpha f_i(x) + u_i)$$

for all (x, u) near $(\overline{x}, \overline{u})$ for each $i \in A(\overline{x})$. Consequently, $\nabla^2_{x,u} L(x, u, \alpha)$ exists near $(\overline{x}, \overline{u})$ and is given by

$$\nabla^2_{x,u} L(x,u,\alpha) = \begin{bmatrix} \nabla_{x,x} L(x,u,\alpha) & f'_E(x)^T & f'_A(x) & 0 \\ f'_E(x) & 0 & 0 & 0 \\ f'_A(x) & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\alpha} I_N \end{bmatrix}.$$

Here we have reordered the components of the vector u into the multipliers associated with the equality constraints u_E with $E = \{s+1, \ldots, m\}$, the multipliers associated with the active inequality constraints u_A with $A = A(\overline{x})$, and the multipliers associated with the inactive inequality constraints u_N with $N = \{1, \ldots, s\} \backslash A(x)$. Also for any matrix $M \subset \mathbb{R}^{m \times n}$ and index set $J \subset \{1, \ldots, m\}$, M_J represents that matrix whose rows are those of M with index in J. Finally,

$$\begin{aligned} \nabla_{xx} L(x,u,\alpha) &= \nabla^2 f_0(x) + \sum_{i \in A} \alpha \nabla f_i(x) \nabla f_i(x)^T + (\alpha f_i(x) + u_i) \nabla^2 f_i(x) \\ &\quad + \sum_{i=s+1}^{m} \alpha \nabla f_i(x) \nabla f_i(x)^T + (\alpha f_i(x) + u_i) \nabla^2 f_i(x) \\ &= \nabla^2_{xx} L(x,u) + \alpha \sum_{i \in A \cup \{s+1,\ldots,m\}} \nabla f_i(x) \nabla f_i(x)^T + f_i(x) \nabla^2 f_i(x) \end{aligned}$$

In order to establish the nonsingularity of $\nabla^2 L(x, u, \alpha)$ we need the following three facts from linear algebra whose proof are left as an exercise.

LEMMA 9.2.1 *Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times n}$ and $D \in \mathbb{R}^{n \times n}$. If*

1. *the rows of D are linearly independent,*
2. *$Dx = 0$, $x \neq 0 \Longrightarrow x^T B x > 0$, and*
3. *$\mu \geq 0$,*

then the matrix $\begin{bmatrix} B + \mu A^T A & D^T \\ D & 0 \end{bmatrix}$ *is nonsingular*

THEOREM 9.2.2 *[Finsler's Theorem] Let $B, C \in \mathbb{R}^{n \times n}$ with C positive semi-definite. Then $x^T B x > 0$ for every $x \in \mathbb{R}^n$, $x \neq 0$ such that $x^T C x = 0$ if and only if $B + \mu C$ is positive definite for all $\mu \geq \overline{\mu}$ for some $\overline{\mu}$.*

THEOREM 9.2.3 *[Debreu's Theorem] Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times n}$. Then $x^T B x > 0$ for every $x \in \mathbb{R}^n$ with $x \neq 0$ such that $Ax = 0$ if and only if $B + \mu A^T A$ is positive definite for all $\mu \geq \overline{\mu}$ for some $\overline{\mu}$.*

The following result is an easy consequence of these linear algebraic results.

THEOREM 9.2.4 *(The positive definiteness of $\nabla^2_{xx}L(\overline{x},\overline{u},\alpha)$ and the nonsingularity of $\nabla^2 L(\overline{x},\overline{u},\alpha)$)*

1. *Let $(\overline{x},\overline{u})$ be a Kuhn-Tucker point for $\mathcal{P}$ and suppose that*

   (a) *(Strict Complementary Slackness)*

   $$\overline{u}_i > 0 \text{ whenever } f_i(x) < 0 \quad i = 1, \ldots, s.$$

   and

   (b) *(Second–Order Sufficiency)*

   $$\nabla f_i(\overline{x})^T d = 0 \quad i \in A(\overline{x}) \cup \{s+1, \ldots, m\} \Longrightarrow d^T \nabla_{11} L(\overline{x},\overline{u}) d > 0.$$

   Then $\nabla_{xx} L(\overline{x},\overline{u},\alpha)$ is positive definite for all $\alpha \geq \overline{\alpha}$ for some $\overline{\alpha} > 0$.

2. *If in addition to the hypotheses in (1) we assume that*

   (c) *(The LI Condition) the gradients $\{\nabla f_i(\overline{x}) : i \in A(\overline{x}) \cup \{s+1, \ldots, m\}\}$ are linearly independent ,*

   Then $\nabla^2 L(\overline{x},\overline{u},\alpha)$ is non-singular for all $\alpha > 0$.

PROOF: (i) We have that

$$\nabla^2_{xx} L(\overline{x},\overline{u},\alpha) = \nabla^2_{xx} L(\overline{x},\overline{u}) + \alpha f_I'(\overline{x}) f_I'(\overline{x})^T$$

where $I = A(\overline{x}) \cup \{s+1, \ldots, m)$. Consequently, the result follows from Debreu's Theorem 9.2.3.

(ii) This just follows from Lemma 9.2.1. ■

We now formally state the method of multipliers.

9.2.1 Algorithm: The Method of Multipliers

Let $(x^0, u^0, \alpha^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+$ stop if $|\nabla L(x^i, u^i, \alpha^i)| < \epsilon$. Having (x^i, u^i, α^i) determine $(x^{i+1}, u^{i+1}, \alpha^{i+1})$ as follows

1. Let x^{i+1} solve

   $$L(x^{i+1}, u^i, \alpha_i) = \min_{x \in \mathbb{R}^n} L(x, u^i, \alpha^i)$$

   or

   $$\nabla_x L(x^{i+1}, u^i, \alpha^i) = 0.$$

2. Set $u^{i+1} = u^i + \alpha_i \nabla_u L(x^{i+1}, u^i, \alpha^i)$ or equivalently

   $$u_j^{i+1} = (\alpha^i f_j(x^{i+1}) + u^i)_+ \quad \text{for } j = 1, \ldots, s$$

   and

   $$u_j^{i+1} = (\alpha^i f_j(x^{i+1}) + u^i) \quad \text{for } j = s+1, \ldots, m.$$

3. Set

$$\alpha^{i+1} := \begin{cases} \alpha^i & \text{if } \|u^{i+1} - u^i\|_\infty \leq \frac{1}{4}\|u^i - u^{i-1}\|_\infty \\ 10\alpha^i & \text{else} \end{cases}$$

In the following theorem we provide a sample of the type of convergence result that can be obtained for this method.

THEOREM 9.2.5 *Let the assumptions (a), (b), and (c) of Theorem 9.2.4 hold and let* $\alpha \geq \overline{\alpha}$. *Let* f_0 *and* f *be* C^2 *near the Kuhn-Tucker point* $\overline{x}$. *Then for* α *sufficiently large, but finite, there is an open neighborhood* V_α *of* $\overline{u}$ *such that for* $u^0 \in V_\alpha$ *there is an* x^0 *such that* $\nabla_x L(x^0, u^0, \alpha) = 0$ *and the iterates* (x^i, u^i) *generated by algorithm 9.2.1 exist and converge to* $(\overline{x}, \overline{u})$ *at the linear root rate*

$$\|x^i - \overline{x}, u^i - \overline{u}\| \leq \delta(\epsilon/\alpha)^i$$

for some positive constants ϵ *and* δ.