Let $\mathcal{F}_t$ denote the information available up to time t, Y_t denote the dependent variable, and $\boldsymbol{x}_t$ be a column vector of explanatory variables. Typically, $\boldsymbol{x}_t$ includes lagged values of the dependent variable. The pth order ARCH model of Engle (1982) can be stated as

$$Y_t = \boldsymbol{x}_t^T\boldsymbol{\beta} + \epsilon_t, E(\epsilon_t|\mathcal{F}_{t-1}) = 0, \text{ and } h_t = \alpha_0 + \psi_1\epsilon_{t-1}^2 + \ldots + \psi_p\epsilon_{t-p}^2, \quad (4.91)$$

where $h_t = var(Y_t|\mathcal{F}_{t-1})$. Note that (4.91) provides a specification not only for the mean but also for the variance. If ψ_j in (4.91) is negative for some j ($j = 1, \ldots, p$), then a large value for ϵ_{t-j} would result in a negative variance for $Y_t|\mathcal{F}_{t-1}$. Therefore, the admissible range for $\psi_1, \ldots, \psi_p$ is $\{\psi_1 \geqslant 0, \ldots, \psi_p \geqslant 0\}$. In the context of this example, the following question is important: "is the variance, h_t, constant over time?" As a general rule, this testing problem is formulated as

$$H_0 : \psi_1 = \ldots = \psi_p = 0 \quad \text{against} \quad H_1 : \psi_1 \geqslant 0, \ldots, \psi_p \geqslant 0. \quad (4.92)$$

In this data example, we consider an ARCH model for an overall rate of return from stocks. Let $Y_t = \log(I_t/I_{t-1})$ where I_t is the *All Ordinaries Index (Australia)* on day t. This index is a weighted average of the prices of selected shares in Australia; it corresponds to the *Dow Jones Index* in the United States. The data that we used are for the period 5/January/1984 to 29/November/1985. This provides a total of 484 observations; see Silvapulle and Silvapulle (1995) for the data. For illustrative purposes, we chose $p = 3$; the method would be applicable for other values of p. The specific model that we consider is

$$\boldsymbol{x}_t^T\boldsymbol{\beta} = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_4 Y_{t-4} \text{ and } h_t = \alpha_0 + \psi_1\epsilon_{t-1}^2 + \psi_2\epsilon_{t-2}^2 + \psi_3\epsilon_{t-3}^2. \quad (4.93)$$

Let $\boldsymbol{\theta} = (\boldsymbol{\lambda} : \boldsymbol{\psi})$ where $\boldsymbol{\lambda} = (\beta_0, \ldots, \beta_4, \alpha_0)^T$ and $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)^T$. The null and alternative hypotheses are

$$H_0 : \boldsymbol{\psi} = \mathbf{0} \text{ and } H_1 : \boldsymbol{\psi} \geqslant \mathbf{0}.$$

Let $\boldsymbol{S}(\boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta})L(\boldsymbol{\theta})$ where $L(\boldsymbol{\theta}) = -(1/2)\sum\{\log(2\pi h_t) + \epsilon_t^2/h_t\}$; thus $L(\boldsymbol{\theta})$ is the loglikelihood if we were to assume that $Y_t|\mathcal{F}_{t-1} \sim N(\boldsymbol{x}_t^T\boldsymbol{\beta}, h_t)$. Now, let us compute the local score-type test statistic S_L in (4.87). First, we need to estimate the nuisance parameter under the null hypothesis. Therefore, we solve the equation $\boldsymbol{S}_{n\lambda}(\bar{\boldsymbol{\theta}}) = \mathbf{0}$ where $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\lambda}} : \mathbf{0})$. Under the null hypothesis, the error variance is constant, and therefore the null model is estimated by ordinary least squares. Thus we have that $\bar{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ and $\bar{\alpha}_0 = n^{-1}\sum(Y_t - \boldsymbol{x}_t^T\bar{\boldsymbol{\beta}})^2$. Now, $\bar{\lambda} = (\bar{\boldsymbol{\beta}} : \bar{\alpha}_0)$. Note that to apply the local score-type test statistic S_L, this is the only estimation required despite the fact that the ARCH model itself is rather involved. The next step is to compute $\bar{\boldsymbol{G}}$, $\bar{\boldsymbol{V}}$, $\bar{\boldsymbol{Z}}$, $\boldsymbol{U}$, and $\bar{\boldsymbol{A}}_{\psi\psi}$. Closed form expressions for these quantities are obtained directly from the closed-form available for $\boldsymbol{S}(\boldsymbol{\theta})$; see Silvapulle and Silvapulle (1995) for these formulas.

The computed value of $\boldsymbol{U}$ is $(3.56, 1.30, 2.20)^T$. Because every component of $\boldsymbol{U}$ is positive, $\min\{(\boldsymbol{U} - \boldsymbol{a})^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}(\boldsymbol{U} - \boldsymbol{a}) : \boldsymbol{a} \geqslant \mathbf{0}\} = 0$ and hence

$$S_L = \boldsymbol{U}^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}\boldsymbol{U} - \min\{(\boldsymbol{U} - \boldsymbol{a})^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}(\boldsymbol{U} - \boldsymbol{a}) : \boldsymbol{a} \geqslant \mathbf{0}\} = \boldsymbol{U}^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}\boldsymbol{U} = 4.36.$$

Let

$$\zeta(\boldsymbol{\lambda}) = [\sum_{i=1}^{3} w_i\{3, \boldsymbol{A}_{\psi\psi}(\boldsymbol{\theta})\}\text{pr}(\chi_j^2 \geqslant 4.36)]_{\boldsymbol{\theta}=(\boldsymbol{\lambda}:\mathbf{0})}.$$

Using the closed forms given in the last chapter for the chi-bar square weights in this expression (see (3.25)), we can compute $sup_\lambda\zeta(\boldsymbol{\lambda})$. However, before we do so, let us apply a bounds test. It follows from (3.23) that

$$\zeta(\boldsymbol{\lambda}) \leq 0.5\{\text{pr}(\chi_2^2 \geqslant 4.36) + \text{pr}(\chi_3^2 \geqslant 4.36)\} = 0.17.$$

This upper bound is too high to reject H_0. Therefore, we computed $p_{sup} = sup_\lambda\zeta(\boldsymbol{\lambda})$ and $\hat{p} = \zeta(\bar{\boldsymbol{\lambda}})$. The computed values are 0.10 and 0.085, respectively. Thus, it appears that there may be some evidence in favor of ARCH effect, but it is weak.

For testing $H_0 : \boldsymbol{\psi} = \mathbf{0}$ against $H_1 : \boldsymbol{\psi} \geqslant \mathbf{0}$, it is also valid to apply an unrestricted test ignoring the inequality constraints in H_1. To this end, let the testing problem be

$$H_0 : \boldsymbol{\psi} = \mathbf{0} \quad \text{vs} \quad H_2 : \boldsymbol{\psi} \neq \mathbf{0}.$$

Even though the ARCH model would not be well-defined with a negative value for any of the components of $\boldsymbol{\psi}$, it is valid to apply a score-type test of $H_0 : \boldsymbol{\psi} = \mathbf{0}$ against $H_2 : \boldsymbol{\psi} \neq \mathbf{0}$; in fact, this is the procedure that is found in much of the econometrics literature. The only information the score-type statistic $\boldsymbol{U}^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}\boldsymbol{U}$ uses is that it is approximately $N(0, 1)$ under H_0; if H_0 is not true then it is expected to be large. The sample value of this is

$$\boldsymbol{U}^T\bar{\boldsymbol{A}}_{\psi\psi}^{-1}\boldsymbol{U} = 4.36;$$

and its large sample p-value is $\text{pr}(\chi_3^2 \geqslant 4.36) = 0.22$. As expected, because the components of $\boldsymbol{U}$ turned out to be all positive, the statistics for the one-sided and two-sided tests have the same numerical value and the latter has a larger p-value.

4.7 APPROXIMATING CONES AND TANGENT CONES

Most of the results in the previous sections have generalizations to the case when the parameter space takes more general forms. Typically, such parameter spaces arise because the constraints are nonlinear in the parameter. It turns out that the main first-order results, namely the asymptotic distribution of an estimator, the asymptotic null distribution of a test statistic, and the asymptotic local power of a test are unchanged when the null and alternative parameter spaces are replaced by cones that approximate them at the true value of the parameter provided that the boundaries of the parameter spaces are sufficiently smooth. The objective of this section is to introduce the basics that are required to study the asymptotics when the parameter spaces are neither linear spaces nor cones; for example, the parameter space may have boundaries that are curved or the slope of the boundary may be discontinuous (i.e., kinked).

As an example, let $\Omega = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 \geqslant g_1(\theta_1) \text{ and } \theta_2 \geqslant g_2(\theta_1)\}$ where $g_1(\theta_1) = \theta_1^2 + \theta_1$ and $g_2(\theta_1) = \theta_1^2 - \theta_1$. In Fig. 4.1, O is the origin; the two

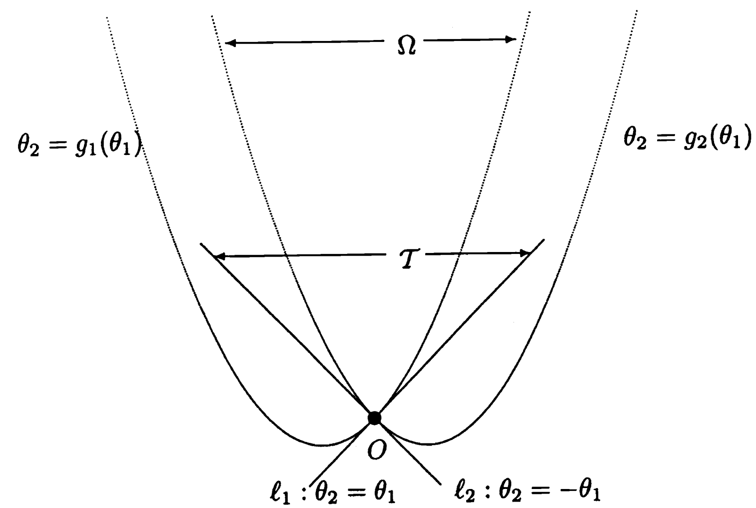

Fig. 4.1 The *cone of tangents*, $\mathcal{T}$, of Ω at O.

parabolas, $\theta_2 = g_1(\theta_1)$ and $\theta_2 = g_2(\theta_1)$, are also shown. Further, Ω is the region bounded by these two parabolas, as shown.

Let ℓ_1 and ℓ_2 be the tangents at O to the two smooth curves that form the boundary of Ω; thus the equations for ℓ_1 and ℓ_2 are $\theta_2 = \theta_1$ and $\theta_2 = -\theta_1$, respectively. Let $\mathcal{T} = \{\boldsymbol{\theta} : \theta_2 \geqslant \theta_1 \text{ and } \theta_2 \geqslant -\theta_1\}$. Locally at the origin, Ω is close to the cone $\mathcal{T}$. In this example, $\mathcal{T}$ is known as the *linearizing cone* of Ω at O, because it is obtained by linearizing the boundaries defining the set Ω at O. The cone is also called the *contingent cone* and *cone of tangents* of Ω at O. The formal definitions of these terms will be introduced in the next subsection.

Let $\boldsymbol{\theta}_0(\neq \mathbf{0})$ be a given fixed point in $\mathbb{R}^p$. Now, let us move $\{\Omega, \ell_1, \ell_2, \mathcal{T}\}$ by $\boldsymbol{\theta}_0$, and let the new objects be denoted by $\{\Theta, \ell_1', \ell_2', \mathcal{A}\}$ (see Fig. 4.2). Therefore, $\Theta = \boldsymbol{\theta}_0 + \Omega, \mathcal{A} = \boldsymbol{\theta}_0 + \mathcal{T}$, ℓ_1' and ℓ_2' are the tangents to the boundaries of Θ at $\boldsymbol{\theta}_0$, and $\mathcal{A}$ is the cone formed by ℓ_1' and ℓ_2'. Then $\mathcal{A}$ is close to Θ at $\boldsymbol{\theta}_0$, in the same way as $\mathcal{T}$ is close to Ω at O. The cone $\mathcal{A}$ with vertex at $\boldsymbol{\theta}_0$ is called the *approximating cone* of Θ at $\boldsymbol{\theta}_0$. An important result relating to the approximating cone is that if Θ is the parameter space and $\boldsymbol{\theta}_0$ is the true value then Θ can be replaced by its approximating cone $\mathcal{A}$ at $\boldsymbol{\theta}_0$ for obtaining the asymptotic null distribution of *LRT* and other first order asymptotic results when $\boldsymbol{\theta}_0$ is the true value.

Example 4.7.1 *Inference on normal mean when the parameter space is not a cone.*

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be *iid* as the bivariate normal distribution, $N(\boldsymbol{\theta}, \boldsymbol{I})$, where $\boldsymbol{\theta} \in \mathbb{R}^2$ and let $\ell(\boldsymbol{\theta})$ denote the kernel of the loglikelihood $-(n/2)\|\bar{\boldsymbol{Y}} - \boldsymbol{\theta}\|^2$. Let the null and alternative hypotheses be

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \text{and} \qquad H_1 : \boldsymbol{\theta} \in \Theta,$$

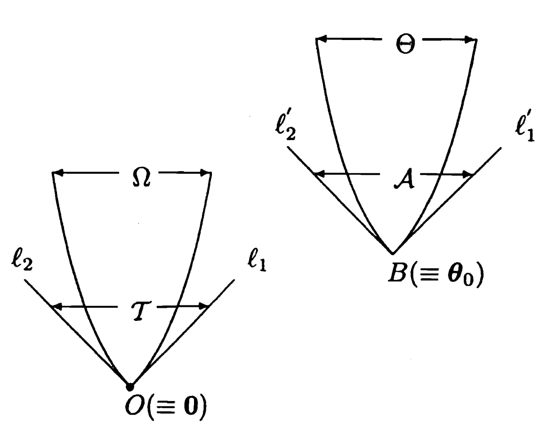

Fig. 4.2 The *cone of tangents*, $\mathcal{T}$, and the *approximating cone*, $\mathcal{A}$, of Θ at B.

respectively, where $\boldsymbol{\theta}_0 = (1,1)^T$, $\Theta = \boldsymbol{\theta}_0 + \Omega$ and Ω is as in Fig. 4.2. Let $\hat{\boldsymbol{\theta}}_\Theta$, $\hat{\boldsymbol{\theta}}_\Omega$ and $\hat{\boldsymbol{\theta}}_{\mathcal{T}}$ denote the *mle* over Θ, Ω, and $\mathcal{T}$, respectively. Then

$$LRT = 2\{\ell(\hat{\boldsymbol{\theta}}_\Theta) - \ell(\boldsymbol{\theta}_0)\} = n\{\|\bar{\boldsymbol{Y}} - \boldsymbol{\theta}_0\|^2 - \|\bar{\boldsymbol{Y}} - \Theta\|^2\}$$

where $\|\bar{\boldsymbol{Y}} - \Theta\|^2 = \inf_{\boldsymbol{\theta}\in\Theta}(\bar{\boldsymbol{Y}} - \boldsymbol{\theta})^T(\bar{\boldsymbol{Y}} - \boldsymbol{\theta})$ is the squared distance between $\bar{\boldsymbol{Y}}$ and Θ. Suppose that the true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}_0$. Now, the results to be presented later in this section show that the asymptotic distribution of the constrained *mle* $\hat{\boldsymbol{\theta}}_\Theta$ and that of the *LRT* of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $\boldsymbol{\theta} \in \Theta$ remain the same even if the constrained parameter space Θ is replaced by $\mathcal{A}$, the approximating cone of Θ at $\boldsymbol{\theta}_0$. More specifically,

$$\begin{aligned} LRT &= n\{\|\bar{\boldsymbol{Y}} - \boldsymbol{\theta}_0\|^2 - \|\bar{\boldsymbol{Y}} - \Theta\|^2\} \\ &= n\{\|\bar{\boldsymbol{Y}} - \boldsymbol{\theta}_0\|^2 - \|\bar{\boldsymbol{Y}} - \mathcal{A}\|^2\} + o_p(1), \qquad (4.94) \\ \text{and} \qquad n^{1/2}&(\hat{\boldsymbol{\theta}}_\Theta - \hat{\boldsymbol{\theta}}_{\mathcal{A}}) = o_p(1). \end{aligned}$$

Further, the constrained estimator $\hat{\boldsymbol{\theta}}_\Theta$ and the projection of the unconstrained estimator $\hat{\boldsymbol{\theta}}$ onto the approximating cone have the same asymptotic distributions in the following sense:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_\Theta - \boldsymbol{\theta}_0) \xrightarrow{d} \Pi(\boldsymbol{Z}, \mathcal{T}) \quad \text{and} \quad \sqrt{n}\{\Pi(\hat{\boldsymbol{\theta}}, \mathcal{A}) - \boldsymbol{\theta}_0\} \xrightarrow{d} \Pi(\boldsymbol{Z}, \mathcal{T}),$$

where $\boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{I})$. ■

It is worth noting that the process of replacing Θ by $\mathcal{A}$ is equivalent to replacing the nonlinear constraints defining Θ by their linear approximations at the null value. In the following subsections, we will extend these ideas to more general settings. To do so, one important requirement is that the parameter space must satisfy certain regularity conditions. These conditions must be satisfied irrespective of the properties of the likelihood function and other stochastic conditions. These are discussed in the next subsection.

Before we introduce regularity conditions relating to parameter spaces, let us make a remark regarding our terminology. Recall that a cone is defined to have vertex at the origin unless the contrary is clear (see Appendix 1). According to our definition, the approximating cone does not necessarily have its vertex at the origin. Some authors refer to the $\mathcal{T}$ in Fig. 4.2 as the approximating cone of Θ at $\boldsymbol{\theta}_0$. We shall deviate from this terminology because it would be helpful to identify $\mathcal{T}$ and $\mathcal{A}$ separately using different terms; we shall refer to them as tangent cone of Θ at $\boldsymbol{\theta}_0$ and approximating cone of Θ at $\boldsymbol{\theta}_0$, respectively.

4.7.1 Chernoff Regularity

Let $\Theta \subset \mathbb{R}^p$ and $\boldsymbol{\theta}_0 \in \Theta$. A vector $\boldsymbol{w} \in \mathbb{R}^p$ is said to be a *tangent* to Θ at $\boldsymbol{\theta}_o$ if there exists a sequence $\{\boldsymbol{\theta}_n\}$ in Θ and a sequence of positive numbers $\{t_n\}$ converging to zero such that $t_n^{-1}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \to \boldsymbol{w}$ as $n \to \infty$. This essentially means that $\boldsymbol{w} \in \mathbb{R}^p$ is a tangent to Θ at $\boldsymbol{\theta}_0$ if either $\boldsymbol{w} = \mathbf{0}$ or there exists a sequence of points in Θ converging to $\boldsymbol{\theta}_0$ such that the direction of the vector from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_n$ converges to the direction of $\boldsymbol{w}$. Therefore, a tangent at $\boldsymbol{\theta}_o$ is a vector rather than the set of points on a line passing through $\boldsymbol{\theta}_0$. This definition of a tangent is more general than the usual one where it is seen as a line that touches a smooth curve at a point. The following examples illustrate this generalized notion of a tangent.

Example 4.7.2 *Illustration of a generalized notion of tangents*

1. For $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 = \theta_1^2\}$, the tangents at $\mathbf{0}$ are $k(1,0)$ and $k(-1,0)$ where $k \geqslant 0$ (see Fig. 4.3). We do not view the line $\theta_2 = 0$ as a tangent.

2. For $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 = \theta_1^2$ and $\theta_1 \geqslant 0\}$ the only tangent at $\mathbf{0}$ is $k(1,0)$ where $k \geqslant 0$; note that $(-1,0)$ is not a tangent at $\mathbf{0}$.

3. For $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 \geqslant \theta_1^2\}$, the tangents at $\mathbf{0}$ are (θ_1, θ_2) where $\theta_2 \geqslant 0$ and $\theta_1 \in \mathbb{R}$ (see Fig. 4.4). Thus, the tangents to the curve $\theta_2 = \theta_1^2$, and those to the area bounded by the curve are different.

4. Let $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 = \sin(1/\theta_1)$ and $\theta_1 > 0\} \cup \{\mathbf{0}\}$. The tangents to Θ at $\mathbf{0}$ are $k\boldsymbol{\theta}$ where $\boldsymbol{\theta} = (\theta_1, \theta_2), \theta_1 > 0$ and $k \geqslant 0$. Thus, even though $(d\theta_2/d\theta_1)$ does not exist at the origin, the set $\{\boldsymbol{\theta} : \theta_2 = \sin(1/\theta_1)\}$ has an infinite number of tangents at the same point. ■

The *cone of tangents,* denoted $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$, of Θ at $\boldsymbol{\theta}_0$ is defined as the set of all tangents to Θ at $\boldsymbol{\theta}_0$. This cone is also called the *contingent cone, Boulingard tangent cone,* and the *ordinary tangent cone.* Note that $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$ has vertex at the origin irrespective of whether or not $\boldsymbol{\theta}_0$ is the origin. A tangent $\boldsymbol{v}$ is said to be *derivable* if there exists a function $\boldsymbol{f} : [0, \epsilon] \to \Theta$ for some $\epsilon > 0$ such that $\boldsymbol{f}(0) = \boldsymbol{\theta}_0$ and $(d/dt+)\boldsymbol{f}(0) = \boldsymbol{v}$. Intuitively, this means that there is a smooth curve in Θ with one end at $\boldsymbol{\theta}_0$ such that the slope of the curve at $\boldsymbol{\theta}_0$ is parallel to $\boldsymbol{v}$. The set of all derivable tangents to Θ at $\boldsymbol{\theta}_0$ is called the *derived tangent cone* of Θ at $\boldsymbol{\theta}_0$. It is also called the *cone of attainable directions*; see Bazaraa et al (1993).

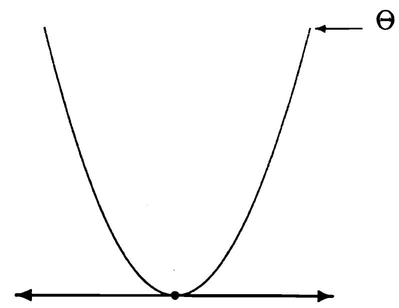

Fig. 4.3 The tangents to $\{\theta_2 = \theta_1^2\}$ at O.

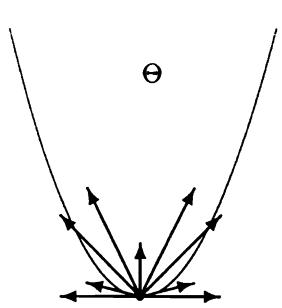

Fig. 4.4 Tangents to $\{\theta_2 \geqslant \theta_1^2\}$ at O.

While it is not essential for us at this stage, let us also state equivalent definitions of the cones just introduced. The derived tangent cone is equal to

$$\{\boldsymbol{v} : \forall t_n \downarrow 0, \exists \boldsymbol{\theta}_n \in \Theta \text{ such that } \boldsymbol{\theta}_n \to \boldsymbol{\theta}_0 \text{ and } t_n^{-1}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \to \boldsymbol{v}\},$$

and the cone of tangents is equal to

$$\{\boldsymbol{v} : \exists t_n \downarrow 0, \exists \boldsymbol{\theta}_n \in \Theta \text{ such that } \boldsymbol{\theta}_n \to \boldsymbol{\theta}_0 \text{ and } t_n^{-1}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \to \boldsymbol{v}\}.$$

It is clear from these definitions that the derived tangent cone is contained in the cone of tangents, because the former must satisfy a condition for all $t_n \downarrow 0$ while the cone of tangents must satisfy the same condition for some $t_n \downarrow 0$.

The following example of (Rockafellar and Wets, 1998, p 199) is helpful to illustrate the difference between the cone of tangents and the derived tangent cone.

Example 4.7.3 *Illustration of a difference between cone of tangents and the derived tangent cone*

Let $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 = \theta_1 \sin(1/\theta_1), \theta_1 > 0\} \cup \{\mathbf{0}\}$. Note that there is no curve that lies in Θ and is smooth at $\mathbf{0}$. Therefore, the only member of the derived tangent cone is $\{\mathbf{0}\}$. By contrast, the straight line $\theta_2 = k\theta_1$ where $|k| \leq 1$, intersects with Θ at a sequence of points $\{\boldsymbol{\theta}_n\}$ and $\|\boldsymbol{\theta}_n\|^{-1}\boldsymbol{\theta}_n \to (1, k)$. Therefore, the cone of tangents is $\{\boldsymbol{\theta} \in \mathbb{R}^2 : |\theta_2| \leq \theta_1, \theta_1 > 0\}$. ■

The foregoing definitions and the example suggest that if the cone of tangents and the derived tangent cone of Θ at $\boldsymbol{\theta}_0$ are not equal, then Θ is likely to be quite irregular near θ_0. A set $\Theta \subset \mathbb{R}^p$ is said to be *Chernoff regular* at $\boldsymbol{\theta}_0$ if the cone of tangents and the derived tangent cone of Θ at $\boldsymbol{\theta}_0$ are equal. If Θ is Chernoff regular at $\boldsymbol{\theta}_0$ then we shall refer to the *cone of tangents* and the *derived tangent cone* as *tangent cone.*

An intuitively simple way of verifying Chernoff regularity of Θ at $\boldsymbol{\theta}_0$ is the following (this is due to Rockafellar and Wets (1998), page 198). Consider the set $\mathcal{A}_t = t(\Theta - \boldsymbol{\theta}_0)$ where $t > 0$. Now take the limit as t increases to ∞ (see Fig. 4.5). Note that $\mathcal{A}_t$ provides a magnified view of Θ near $\boldsymbol{\theta}_0$ because it is like viewing Θ

through a magnifying glass focused at $\boldsymbol{\theta}_0$. As t increases to ∞, the magnification increases to ∞ and if the magnified view of Θ at $\boldsymbol{\theta}_0$ converges to a limit (i.e., if $\mathcal{A}_t$ has a limit) then Θ is Chernoff regular at $\boldsymbol{\theta}_0$, and the limit is the tangent cone of Θ at $\boldsymbol{\theta}_0$.

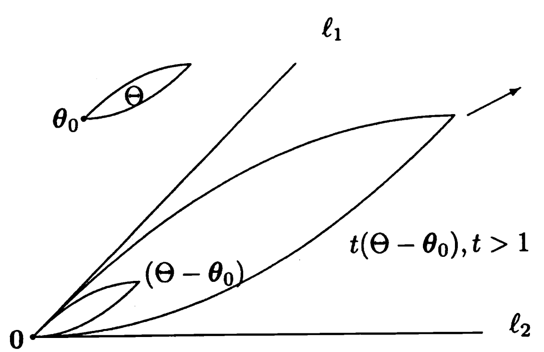

Fig. 4.5 The limit of $t(\Theta - \boldsymbol{\theta}_0)$ as $t \to \infty$ is the cone between the lines ℓ_1 and ℓ_2, and it is equal to the tangent cone, $\mathcal{T}(\Theta; \boldsymbol{\theta}_0)$, of Θ at $\boldsymbol{\theta}_0$.
Reproduced from: *Variational Analysis*, Rockafellar and Wets, Copyright (1998), with permission from Springer, New York.

Let $\Theta \subset \mathbb{R}^p$ and $\boldsymbol{\theta}_0 \in \Theta$. The set Θ is said to be *approximated by a cone $\mathcal{A}$ at $\boldsymbol{\theta}_0$* if

$$\begin{aligned} (a) \quad & d(\boldsymbol{\theta}, \mathcal{A}) = o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|) \text{ for } \boldsymbol{\theta} \in \Theta, \\ \text{and } (b) \quad & d(\boldsymbol{x}, \Theta) = o(\|\boldsymbol{x} - \boldsymbol{\theta}_0\|) \text{ for } \boldsymbol{x} \in \mathcal{A}, \end{aligned} \tag{4.95}$$

where d is the distance between a point and a set defined by

$$d(\boldsymbol{\theta}, A) = \inf_{\boldsymbol{x} \in A} \|\boldsymbol{\theta} - \boldsymbol{x}\|.$$

In this case, $\mathcal{A}$ is called an *approximating cone of Θ at $\boldsymbol{\theta}_0$* and will be denoted by $\mathcal{A}(\Theta_0; \boldsymbol{\theta}_0)$.

This definition was introduced by Chernoff (1954) and it is illustrated in Fig. 4.6 using two different shapes for Θ. In each of the two cases, let Θ be the region bounded by the two curves as shown and O be the point corresponding to $\boldsymbol{\theta}_0$. In the diagram on the right-hand side, A is a point in Θ. Let $\boldsymbol{\theta} = OA$ and B be the projection of A onto the approximating cone $\mathcal{A}$; therefore, $\|AB\| = d(\boldsymbol{\theta}, \mathcal{A})$. Condition (a) of the definition (4.95) says that $\|AB\|/\|OA\| \to 0$ as A converges to O while staying in Θ. In the diagram on the left-hand side, A is a point in $\mathcal{A}$ and let $\boldsymbol{x} = OA$. Let B be the projection of A on to Θ. Condition (b) of the definition (4.95) says that $\|AB\|/\|OA\| \to 0$ as A converges to O while staying in $\mathcal{A}$.

Andrews (1999, page 1358) noted that the foregoing definition of approximating cone can be stated using Hausdorff distance; see also Le Cam (1970, p 819). For two sets A and B, define the *Hausdorff distance* between A and B by

$$h(A, B) = \max\{\sup_{b \in B} d(b, A), \sup_{a \in A} d(a, B)\}.$$

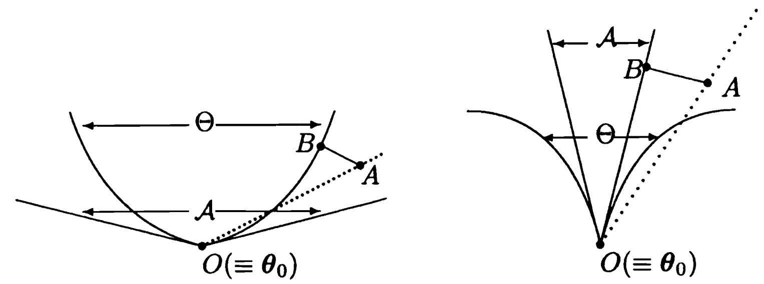

Fig. 4.6 Chernoff's definition of approximating cone $\mathcal{A}$ of Θ at $\boldsymbol{\theta}_0$: $\|AB\| = o(\|OA\|$

Now, the next result provides an equivalent form of definition (4.95).

Proposition 4.7.1 *Let the setting be as in* (4.95) *and* $B_\epsilon = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon\}$, *where* $\epsilon > 0$. *Let $\mathcal{A}$ be a cone with vertex at $\boldsymbol{\theta}_0$. Then $\mathcal{A}$ is an approximating cone of Θ at $\boldsymbol{\theta}_0$ as defined by* (4.95) *if and only if*

$$h\{\Theta \cap B_\epsilon, \mathcal{A} \cap B_\epsilon\} = o(\epsilon) \quad \text{as } \epsilon \searrow 0. \tag{4.96}$$

Proof: For a given $\epsilon > 0$ let $\Theta_\epsilon = \Theta \cap B_\epsilon, \mathcal{A}_\epsilon = \mathcal{A} \cap B_\epsilon, \boldsymbol{\theta}_\epsilon = \arg\max_{\boldsymbol{\theta} \in \Theta_\epsilon} d(\boldsymbol{\theta}, \mathcal{A}_\epsilon)$, and $\boldsymbol{x}_\epsilon = \arg\max_{\boldsymbol{x} \in \mathcal{A}_\epsilon} d(\boldsymbol{x}, \Theta_\epsilon)$. Then, the Hausdorff distance between Θ_ϵ and $\mathcal{A}_\epsilon$ takes the form

$$h(\Theta_\epsilon, \mathcal{A}_\epsilon) = \max\{d(\boldsymbol{\theta}_\epsilon, \mathcal{A}_\epsilon), d(\boldsymbol{x}_\epsilon, \Theta_\epsilon)\}.$$

Now suppose that $\mathcal{A}$ is an approximating cone according to Chernoff's definition (4.95). Then

$$d(\boldsymbol{\theta}_\epsilon, \mathcal{A}_\epsilon) = o(\|\boldsymbol{\theta}_\epsilon - \boldsymbol{\theta}_0\|) = o(\epsilon);$$

the first equality follows by (4.95) and the second follows since $\|\boldsymbol{\theta}_\epsilon - \boldsymbol{\theta}_0\| \leq \epsilon$. Similarly, $d(\boldsymbol{x}_\epsilon, \Theta_\epsilon) = o(\epsilon)$. Therefore, $h(\Theta_\epsilon, \mathcal{A}_\epsilon) = o(\epsilon)$.

Now to prove the converse, suppose that (4.96) is satisfied. Let $\boldsymbol{\theta} \in \Theta$ and $\delta = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$. Then,

$$d(\boldsymbol{\theta}, \mathcal{A}) = d(\boldsymbol{\theta}, \mathcal{A}_\delta) \leq d(\boldsymbol{\theta}_\delta, \mathcal{A}_\delta) = o(\delta);$$

the first equality follows because $\|\Pi(\boldsymbol{\theta}, \mathcal{A}) - \boldsymbol{\theta}_0\| \leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$ and hence $\Pi(\boldsymbol{\theta}, \mathcal{A}) = \Pi(\boldsymbol{\theta}, \mathcal{A})$, the middle inequality follows by the definition of $\boldsymbol{\theta}_\delta$, and the third equality follows by (4.96). ■

For a discussion about Hausdorff distance in statistics, see van der Vaart and Wellner (1996, p 162). Andrews (1999, p 1358) extended the definition of Chernoff (1954) to define a sequence of sets being approximated by a cone; such a generalization may be required/helpful in more general settings. We say that a sequence of sets $\{\Theta_n\}$ is locally approximated at $\boldsymbol{\theta}_0$ by a cone $\mathcal{A}$ if

$$\begin{aligned} (a) \quad & d(\boldsymbol{\theta}_n, \mathcal{A}) = o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|) \text{ for } \boldsymbol{\theta}_n \in \Theta_n \text{ and } \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\| \to 0 \\ \text{and } (b) \quad & d(\boldsymbol{x}_n, \Theta_n) = o(\|\boldsymbol{x}_n - \boldsymbol{\theta}_0\|) \text{ for } \boldsymbol{x}_n \in \mathcal{A} \text{ and } \|\boldsymbol{x}_n - \boldsymbol{\theta}_0\| \to 0. \end{aligned} \tag{4.97}$$

Now the foregoing results for Θ also hold for Θ_n with appropriate modifications.

Throughout this book, parameter spaces will be assumed to be Chernoff regular unless the contrary is made clear; this is a very mild condition as was indicated earlier. The following important result, due to Geyer (1994, Theorem 2.1), establishes the link between the notion of approximating cone introduced by Chernoff and several tangent cones that have been known in the mathematics literature.

Proposition 4.7.2 *(Geyer 1994). The cone $\mathcal{K}$ is an approximating cone of Θ at θ_0 if and only if the derived tangent cone and the cone of tangents are equal and*

$$\text{Closure of } \mathcal{K} = \theta_0 + \mathcal{T}(\Theta;\theta_0) = \theta_0 + \text{Derived tangent cone of } \Theta \text{ at } \theta_0.$$

■

This essentially says that Θ has an approximating cone at θ_0 if an only if Θ is Chernoff regular at θ_0.

The following result provides a convenient way of obtaining the approximating cone of Θ at θ_0 when Θ is defined by a set of nonlinear equality and inequality constraints, in many cases; the proof is given in the appendix to this chapter.

Proposition 4.7.3 *Suppose that Ω is open and let*

$$\Theta = \{\theta \in \Omega : h_1(\theta) = \ldots = h_\ell(\theta) = 0, h_{\ell+1}(\theta) \geqslant 0, \ldots, h_k(\theta) \geqslant 0\},$$

where $h_1, \ldots, h_k$ are continuously differentiable. Let θ_0 be a point in Θ, and let $a_i = (\partial/\partial\theta)h_i(\theta_0)$ for $i = 1, \ldots, k$, and $J(\theta_0) = \{i : h_i(\theta_0) = 0$ and $\ell + 1 \leq i \leq k\}$. Assume that the following condition, known as the Mangasarian-Fromowitz constraint qualification (MF-CQ), is satisfied at θ_0: There exists a nonzero $b \in \mathbb{R}^p$ such that $a_1^T b = \ldots = a_\ell^T b = 0$, $a_1, \ldots, a_\ell$ are linearly independent and $a_i^T b > 0$ for $i \in J(\theta_0)$. Then $\mathcal{A}(\Theta;\theta_0)$ is equal to

$$\{\theta \in \mathbb{R}^p : a_i^T(\theta - \theta_0) = 0 \text{ for } i = 1, \ldots, \ell; a_i^T(\theta - \theta_0) \geqslant 0 \text{ for } i \in J(\theta_0)\}.$$

The foregoing result shows that if MF-CQ is satisfied then the approximating cone at θ_0 is obtained by substituting the first order approximation (i.e., linear approximation at θ_0), $\{h_i(\theta_0)+(\theta-\theta_0)^T(\partial/\partial\theta)h_i(\theta_0)\}$, of $h_i(\theta)$ at θ_0 if $h_i(\theta_0) = 0$; if $h_s(\theta_0) > 0$ (i.e., $h_s(\theta_0) \geq 0$ is inactive) then the constraint $h_s(\theta_0) \geqslant 0$ does not play a part in determining the approximating cone at θ_0 $(s = \ell + 1, \ldots, k)$.

The approximating cone obtained by applying the result in the previous proposition is a polyhedral. Thus, if the approximating cone is not a polyhedral then this method of obtaining the approximating cone would not be applicable in its current form. For example, in Fig. 4.1 if $\Theta = A \cup B$ where A is the region bounded by $\theta_2 = g_2(\theta_1)$ and $\theta_2 = \theta_1$, and B is the region bounded by $\theta_2 = g_1(\theta_1)$ and $\theta_2 = -\theta_1$, then the approximating cone of Θ at O is not convex. Therefore, for this Θ, the method in Proposition 4.7.3 is not directly applicable; in any case, this would be obvious because Θ is not of the form required by Proposition 4.7.3. It can be seen that A and B are

of the form required by Proposition 4.7.3 and $\mathcal{A}(A \cup B; \mathbf{0}) = \mathcal{A}(A;\mathbf{0}) \cup \mathcal{A}(B;\mathbf{0})$; now the method based on MF-CQ can be applied to A and B separately to obtain $\mathcal{A}(A \cup B;\mathbf{0})$. More generally, the foregoing method can be applied to a parameter space that is equal to the union of sets each of which satisfies the conditions of Proposition 4.7.3.

Example 4.7.4 *Tangent Cones*

(1) Let $\Theta = \{\theta \in \mathbb{R}^2 : \theta_2 \geqslant (1/2)\theta_1, \theta_2 \leq 2\theta_1\}$ (see Fig. 4.7). Then, we have the following:
$\mathcal{T}(\Theta_0;\theta_0) = \Theta$, where $\theta_0 = (0,0)^T$.
$\mathcal{T}(\Theta;\theta_a) = \{x \in \mathbb{R}^2 : x_2 \geqslant (1/2)x_1\}$, where $\theta_a = (2,1)^T$.
$\mathcal{T}(\Theta;\theta_b) = \mathbb{R}^2$, where $\theta_b = (2,2)^T$.
$\mathcal{T}(\Theta;\theta_c) = \{x \in \mathbb{R}^2 : x_2 \leq 2x_1\}$, where $\theta_c = (1,2)^T$.

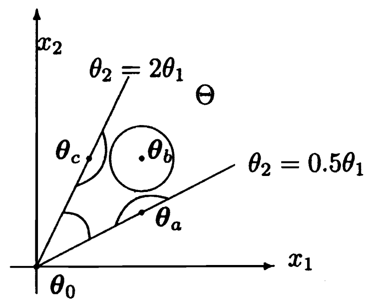

Fig. 4.7 The tangent cones of $\Theta = \{\theta : \theta_2 \leq 2\theta_1, \theta_2 \geqslant 0.5\theta_1\}$ at $\theta_0, \theta_a, \theta_b$ and θ_c

Note that at θ_a, the inequality $\theta_2 \leq 2\theta_1$ holds strictly (i.e., it is not active) and therefore this inequality constraint can be ignored for determining the tangent cone at θ_a. At θ_b, both inequalities are satisfied strictly, and therefore both constraints can be ignored for determining the tangent cone. Since there are no other active constraints, the tangent cone is the full space $\mathbb{R}^2$. At θ_c, the inequality $\theta_2 \geqslant (1/2)\theta_1$ holds strictly and hence it can be ignored for determining the tangent cone; the rest of the arguments are similar to those for θ_a. ■

(2) Let the matrices R_1 and R_2 be given and let $\Theta = \{\theta \in \mathbb{R}^p : R_1\theta \geqslant \mathbf{0}, R_2\theta \geqslant \mathbf{0}\}$. Then, by Proposition 4.7.3 we have the following:

$$\begin{aligned} &R_1\theta_a > \mathbf{0}, R_2\theta_a = \mathbf{0} \Rightarrow \mathcal{T}(\Theta;\theta_a) = \{x \in \mathbb{R}^p : R_2 x \geqslant \mathbf{0}\}.\\ &R_1\theta_0 = \mathbf{0}, R_2\theta_0 = \mathbf{0} \Rightarrow \mathcal{T}(\Theta;\theta_0) = \{x \in \mathbb{R}^p : R_1 x \geqslant \mathbf{0}, R_2 x \geqslant \mathbf{0}\}. \qquad (4.98)\\ &R_1\theta_b > \mathbf{0}, R_2\theta_b > \mathbf{0} \Rightarrow \mathcal{T}(\Theta;\theta_b) = \mathbb{R}^p. \end{aligned}$$

(3) Let $\Theta = \{\theta \in \mathbb{R}^2 : \theta_2 \geqslant -\theta_1^2 + 2\theta_1, \theta_2 \geqslant -\theta_1^2 - 2\theta_1\}$ and $\theta_0 = \mathbf{0}$.

It is easily verified that the MF-CQ is satisfied at θ_0 and therefore the tangent cone is determined by linearizing the constraints at θ_0 as follows. Let the two constraints be expressed as $g_1(\boldsymbol{\theta}) \geqslant 0$ and $g_2(\boldsymbol{\theta}) \geqslant 0$. Then the linearized versions of these constraints at $\boldsymbol{\theta}_0$ are

$$g_1(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T(\partial/\partial\boldsymbol{\theta})g_1(\boldsymbol{\theta}_0) \geqslant 0 \text{ and } g_2(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T(\partial/\partial\boldsymbol{\theta})g_2(\boldsymbol{\theta}_0) \geqslant 0,$$

respectively. These linearized versions are precisely $\theta_2 \geqslant 2\theta_1$ and $\theta_2 \geqslant -2\theta_1$, respectively. Since the two constraints defining Θ are active at $\boldsymbol{\theta}_0$, we have $\mathcal{T}(\Theta; \boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : \theta_2 \geqslant 2\theta_1, \theta_2 \geqslant -2\theta_1\}$. In this particular example, we can sketch the shape of Θ easily and the tangent cone can be written down quickly; the above analytical approach is meant to be instructive.

(4) Let $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_2 \geqslant \theta_1^2, \theta_2 \leq \theta_1\}$. Let $g_1(\boldsymbol{\theta}) = \theta_2 - \theta_1^2$ and $g_2(\boldsymbol{\theta}) = \theta_1 - \theta_2$. Then $\nabla g_1(\boldsymbol{\theta}) = (-2\theta_1, 1)^T$, $\nabla g_2(\boldsymbol{\theta}) = (1, -1)^T$. Now, applying the same technique as in the previous example, we have $\mathcal{T}(\Theta; \boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_1 \geqslant \theta_2 \geqslant 0\}$, and Θ is Chernoff regular at $\mathbf{0}$ (see Fig. 4.8). ■

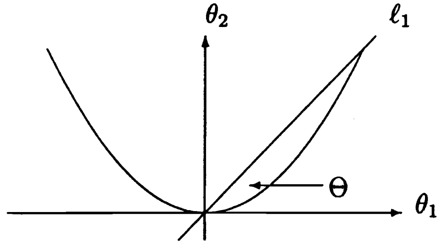

Fig. 4.8 The tangent cone of $\Theta = \{\theta_2 \geqslant \theta_1^2, \theta_2 \leq \theta_1\}$ at the origin is the cone between ℓ_1 and the θ_1 axis.

The next result is useful in asymptotic derivations because it enables us to substitute the approximating cone for the parameter space at the true value even if it (i.e., the true parameter value) is on the boundary which may or may not be smooth.

Proposition 4.7.4 *Let $\Theta \subset \mathbb{R}^p$, $\boldsymbol{V}$ be a positive definite matrix of order $p \times p$ and $\boldsymbol{\theta}_0 \in \Theta$ and $\mathcal{A}$ be an approximating cone of Θ at $\boldsymbol{\theta}_0$. Then*

$$\|\boldsymbol{y} - \Theta\|_{\boldsymbol{V}}^2 - \|\boldsymbol{y} - \mathcal{A}(\Theta; \boldsymbol{\theta}_0\|_{\boldsymbol{V}}^2 = o(\|\boldsymbol{y} - \boldsymbol{\theta}_0\|_{\boldsymbol{V}}^2) \text{ as } \boldsymbol{y} \to \boldsymbol{\theta}_0.$$

The proof is given in the Appendix; in fact, two different proofs are given, one uses Chernoff's definition of an approximating cone and simple geometry and the other uses Hausdorff distance. Fig. 4.9 illustrates this result and it shows two possible shapes of Θ bounded by two curves intersecting at O. Let A be an arbitrary point, and C and D be the points in $\mathcal{A}(\Theta; \boldsymbol{\theta}_0)$ and Θ, respectively, that are closest to A. Now Proposition 4.7.4 says that $|AD^2 - AC^2| = o(OA^2)$.

A useful corollary to the foregoing proposition is the following.

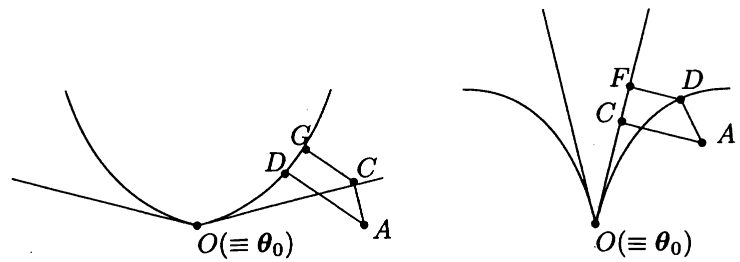

Fig. 4.9 $\|\boldsymbol{y} - \Theta\|^2 - \|\boldsymbol{y} - \mathcal{A}(\Theta; \boldsymbol{\theta}_0)\|^2 = o(\|\boldsymbol{y} - \boldsymbol{\theta}_0\|^2)$ [i.e., $|AD^2 - AC^2| = o(OA^2)$]

Corollary 4.7.5 *Let $\Theta \subset \mathbb{R}^p$, $\boldsymbol{\theta}_0 \in \Theta$, $\boldsymbol{V}$ be a $p \times p$ positive definite matrix and Θ be Chernoff regular at $\boldsymbol{\theta}_0$. Then we have the following:*

1. *If $n^{1/2}(\boldsymbol{T}_n - \boldsymbol{\theta}_0) = O_p(1)$ then*

$$\|\boldsymbol{T}_n - \Theta\|_{\boldsymbol{V}}^2 - \|\boldsymbol{T}_n - \mathcal{A}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2 = o_p(n^{-1}).$$

2. *If $n^{1/2}(\boldsymbol{T}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \boldsymbol{Z}$ then $n\|\boldsymbol{T}_n - \Theta\|_{\boldsymbol{V}}^2 \xrightarrow{d} \|\boldsymbol{Z} - \mathcal{T}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2$*

Proof: Let us first state a lemma.

Lemma 4.7.6 *Let $g : \mathbb{R}^p \to \mathbb{R}$ be such that $g(\boldsymbol{y}) = o(\|\boldsymbol{y} - \boldsymbol{\theta}_0\|^\beta)$ as $\|\boldsymbol{y} - \boldsymbol{\theta}_0\| \to 0$, where $\beta > 0$. Let $\boldsymbol{T}_n$ be a sequence of random p-vectors such that $\boldsymbol{T}_n - \boldsymbol{\theta}_0 = O_p(n^{-\gamma})$ where $\gamma > 0$. Then $g(\boldsymbol{T}_n) = o_p(n^{-\beta\gamma})$.*

Proof: Given $\epsilon > 0$, there exists a δ such that $|g(\boldsymbol{y})/\|\boldsymbol{y} - \boldsymbol{\theta}_0\|^\beta| < \epsilon$ for $\|\boldsymbol{y} - \boldsymbol{\theta}_0\| < \delta$. Therefore,

$$\mathrm{pr}\{|g(\boldsymbol{T}_n)/\|\boldsymbol{T}_n - \boldsymbol{\theta}_0\|^\beta < \epsilon) \geqslant \mathrm{pr}(\|\boldsymbol{T}_n - \boldsymbol{\theta}_0\| < \delta) \to 1 \text{ as } n \to \infty.$$

Thus, $g(\boldsymbol{T}_n)/\|\boldsymbol{T}_n - \boldsymbol{\theta}_0\|^\beta = o_p(1)$. Now,

$$g(\boldsymbol{T}_n) = \|\boldsymbol{T}_n - \boldsymbol{\theta}_0\|^\beta o_p(1) = |n^{-\gamma}O_p(1)|^\beta o_p(1) = o_p(n^{-\beta\gamma}).$$

This completes the proof of the lemma.

Now, the proof of the first part of the corollary follows from the foregoing Lemma and Proposition 4.7.4 by substituting $g(\boldsymbol{y}) = \|\boldsymbol{y} - \Theta\|_{\boldsymbol{V}}^2 - \|\boldsymbol{y} - \mathcal{A}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2$, $\beta = 2$ and $\gamma = 1/2$.

To prove the second part, let $\boldsymbol{Z}_n = n^{1/2}(\boldsymbol{T}_n - \boldsymbol{\theta}_0)$ and note that

$$\begin{aligned} n\|\boldsymbol{T}_n - \Theta\|_{\boldsymbol{V}}^2 &= n\|\boldsymbol{T}_n - \mathcal{A}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2 + o_p(n^{-1}) \\ &= \|\boldsymbol{Z}_n - \mathcal{T}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2 + o_p(1) \xrightarrow{d} \|\boldsymbol{Z} - \mathcal{T}(\Theta; \boldsymbol{\theta}_0)\|_{\boldsymbol{V}}^2 \end{aligned}$$

■

The corollary essentially says that $\mathcal{A}(\Theta;\boldsymbol{\theta}_0)$ can be substituted for Θ in most of the expressions; in particular, if $\boldsymbol{T}_n$ is a $\sqrt{n}$-consistent estimator, $\boldsymbol{\theta}_0$ is the true value and $\boldsymbol{\theta}_0 \in \Theta$ then the squared distance between $\boldsymbol{T}_n$ and Θ is equal to that between $\boldsymbol{T}_n$ and the approximating cone of Θ at $\boldsymbol{\theta}_0$ except for a $o_p(n^{-1})$ term. The corollary can also be extended to the case when Θ_n is a sequence of sets such that $\boldsymbol{\theta}_0 \in \Theta_n$ and $\mathcal{A}(\Theta_n;\boldsymbol{\theta}_0)$ converges to a cone, or Θ_n converges to a cone.

4.8 GENERAL TESTING PROBLEMS

Let $f(y;\boldsymbol{\theta})$ be a density function where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ and $\Theta_0 \subset \Theta_1 \subset \Theta$. Let Y be a random variable with density function $f(y;\boldsymbol{\theta})$. In this section, test of

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\theta} \in \Theta_1 \tag{4.99}$$

is studied when Θ_0, Θ_1 and Θ may have boundary points and the boundaries of these sets may or may not be smooth; for example, the boundary of a cone at its vertex is not smooth. The special case when Θ_0 is a linear space and Θ_1 is a polyhedral was studied in Section 4.3, where the special structure on the parameter spaces enabled us to obtain explicit expressions for the asymptotic null distributions of the test statistics. Now, we consider more general cases and obtain general results. It will be assumed that Θ_0 and Θ_1 have approximating cones at every point in Θ_0; in the terminology introduced in Section 4.7, this is same as saying that Θ_0 and Θ_1 are Chernoff regular at every $\boldsymbol{\theta} \in \Theta_0$. For simplicity, it is assumed that the observations on Y are independently and identically distributed; however, the main results of this section hold, under appropriate regularity conditions, for much more general stochastic processes although such generalizations are not considered.

Let $Y_1, \ldots, Y_n$ be independently and identically distributed as Y. As in the earlier sections, let

$$\ell(\boldsymbol{\theta}) = \sum \log f(Y_i;\boldsymbol{\theta}), \qquad \boldsymbol{S}(\boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta})\ell(\boldsymbol{\theta}), \tag{4.100}$$

$$\text{and} \quad \mathcal{I}(\boldsymbol{\theta}) = E\{(\partial/\partial\boldsymbol{\theta}) \log f(Y;\boldsymbol{\theta})(\partial/\partial\boldsymbol{\theta}^T) \log f(Y;\boldsymbol{\theta})\} \tag{4.101}$$

denote the loglikelihood, the score function, and the information matrix, respectively. Some important fundamental results on the *LRT* for testing problems of the form (4.99) were obtained by Chernoff (1954). The technical details in Chernoff (1954) assumed that Θ was open and that inference was based on likelihood. Chernoff's (1954) results have natural extensions to more general cases although modifications would be required to the regularity conditions and technical details; for example, see Self and Liang (1987), Andrews (1998), Vu and Zhou (1997), Shapiro (1985, 1989, 2000a), Geyer (1994), Silvapulle (1992a, 1992b, 1992c, 1994), and El Barmi (1996) among others. The areas covered include the cases when (i) Θ is not necessarily open (ii) observations are not necessarily independently and identically distributed, and (iii) inference is based on quasi-likelihood, partial likelihood, or general objective functions such as those used in, for example, M-estimation, minimum distance methods, and empirical likelihood.

In the next subsection, we consider the simple case of likelihood ratio test when Θ is open. The simplicity of this case makes it easier to convey the essentials. On the other hand, this case is also of independent interest because many practical problems fall into this category. Then, in subsection 4.8.2, we consider likelihood inference when Θ is not necessarily open and the true value of $\boldsymbol{\theta}$ can be a boundary point of Θ, Θ_0, and Θ_1 simultaneously. The distribution and implementation of likelihood ratio test are virtually the same whether or not $\boldsymbol{\theta}_0$ is a boundary point of Θ provided some conditions are satisfied; essentially all that we require is that it should be possible to approximate the objective function by a quadratic similar to that in (4.2) and (4.3). In this sense, the results in subsection 4.8.1 are not very different from those in subsection 4.8.2. In subsection 4.8.3, we consider inference based on general objective functions rather than just likelihood. Some results obtained in subsections 4.8.1 and 4.8.2 do not necessarily remain the same in this general setting. In subsection 4.8.4, some simple examples in two dimensions are considered to illustrate the theory. In subsection 4.8.5, a general form of the Type B testing problem, test of $\boldsymbol{h}(\boldsymbol{\theta}) \geqslant \boldsymbol{0}$ against $\boldsymbol{h}(\boldsymbol{\theta}) \not\geqslant \boldsymbol{0}$, is studied; this particular type of problems arises in economics, for example, the interest may be to test whether or not a regression function deviates from concavity.

As usual, we adopt the notation $\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}$, and $\bar{\boldsymbol{\theta}}$ for the estimators over Θ, Θ_1, and Θ_0 respectively.

4.8.1 Likelihood Approach When Θ is Open

Assume that Θ is open and that $H_0 : \boldsymbol{\theta} \in \Theta_0$ holds. Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$ in Θ_0. The set of regularity conditions required for this setting are virtually the same as those usually required for establishing that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal and the asymptotic null distribution of the *LRT* of $\boldsymbol{R\theta} = \boldsymbol{0}$ against $\boldsymbol{R\theta} \neq \boldsymbol{0}$ is chi-squared. There are different sets of conditions available for these results; some involve third-order partial derivatives of the log density, some involve only the second-order partial derivatives, and some replace the differentiability by other special types of differentiability (quadratic mean, stochastic, etc.). For simplicity, we shall work with conditions on the third-order derivatives of the log density, but we recognize that our technical conditions can be improved. Let

$$\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T \mathcal{I}_{\boldsymbol{\theta}_0} \boldsymbol{x}. \tag{4.102}$$

Define,

$$W_D = \min_{\boldsymbol{\theta} \in \Theta_0} n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \hat{\mathcal{I}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \min_{\boldsymbol{\theta} \in \Theta_1} n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \hat{\mathcal{I}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{4.103}$$

where $\hat{\mathcal{I}} = \mathcal{I}(\hat{\boldsymbol{\theta}})$; however, in general $\hat{\mathcal{I}}$ can be any consistent estimator of $\mathcal{I}(\boldsymbol{\theta}_0)$. Note that W_D is a measure of $\{dist(\hat{\boldsymbol{\theta}}, H_0) - dist(\hat{\boldsymbol{\theta}}, H_1)\}$, where $dist(\hat{\boldsymbol{\theta}}, H)$ is a measure of distance between $\hat{\boldsymbol{\theta}}$ and the parameter space defined by H. Hence it is a suitable statistic for testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_1 : \boldsymbol{\theta} \in \Theta_1$. As in the case of test against the unrestricted alternative, *LRT* and W_D are asymptotically equivalent. A simple method of obtaining their common asymptotic distribution is given in the