CSCanada
Progress in Applied Mathematics
Vol. 4, No. 2, 2012, pp. [167–181]
DOI: 10.3968/j.pam.1925252820120402.S923

ISSN 1925-251X [Print]
ISSN 1925-2528 [Online]
www.cscanada.net
www.cscanada.org

Constrained Statistical Inference: A Hybrid of Statistical Theory, Projective Geometry and Applied Optimization Techniques

Karelyn A. Davis[a],*

[a] Environmental Health Science and Research Bureau, Health Canada, Ottawa, ON, Canada.

* Corresponding author.
Address: Environmental Health Science and Research Bureau, Health Canada, Ottawa, ON, K1A 0K9, Canada; E-Mail: karelyn.davis@hc-sc.gc.ca

Received: August 18, 2012/ Accepted: October 11, 2012/ Published: October 31, 2012

Abstract: In many data applications, in addition to determining whether a given risk factor affects an outcome, researchers are often interested in whether the factor has an increasing or decreasing effect. For instance, a clinical trial may test which dose provides the minimum effect; a toxicology study may wish to determine the effect of increasing exposure to a harmful contaminant on human health; and an economist may wish to determine an individual's optimal preferences subject to a budget constraint. In such situations, constrained statistical inference is typically used for analysis, as estimation and hypothesis testing incorporate the parameter orderings, or restrictions, in the methodology. Such methods unite statistical theory with elements of projective geometry and optimization algorithms. In many different models, authors have demonstrated constrained techniques lead to more efficient estimates and improved power over unconstrained methods, albeit at the expense of additional computation. In this paper, we review significant advancements made in the field of constrained inference, ranging from early work on isotonic regression for several normal means to recent advances of constraints in Bayesian techniques and mixed models. To illustrate the methods, a new analysis of an environmental study on the health effects in a population of newborns is provided.

Key words: Constrained inference; Convex cone; Cord blood; Gradient projection; Linear regression; Maximum likelihood

Davis, K.A. (2012). Constrained Statistical Inference: A Hybrid of Statistical Theory, Projective Geometry and Applied Optimization Techniques. *Progress in Applied Mathematics, 4*(2), 167–181. Available from http://www.cscanada.net/index.php/pam/article/view/j.pam.1925252820120402.S923 DOI: 10.3968/j.pam.1925252820120402.S923

1. INTRODUCTION

As with many problems in statistical analysis, a parameter space is defined from quantities of interest, and subsequent analysis aims to conduct hypothesis testing to make comparisons. Relationships between parameters often arise naturally from the context of the experiment or observational study. For instance, in clinical trials, a researcher may desire a method to only declare a lower dose to be efficacious if a higher dose is first found to be efficacious; the severity of an infection may decrease over time and such a relationship may differ for two treatment groups; estimation in mixed model analysis must ensure the covariance matrix is positive definite; or a National hockey league owner may be interested in determining if the selection of a player of high rank in the Entry Draft will lead to improved team performance (Peng *et al.* [1], Molenberghs & Verbeke [2], Dawson & Magee [3]).

Such constraints on statistical parameters have important implications for estimation and hypothesis testing. These procedures have been studied under various names in the literature—one-sided testing, isotonic regression, restricted analysis, etc. The choice of a constrained method, while computationally and theoretically more intensive, usually leads to a more efficient method than models which ignore the constraints. The seminal texts by Barlow *et al.* [4], Robertson, Wright and Dykstra [5] and Silvapulle and Sen [6] highlight the many achievements in constrained inference over the past sixty years. In this paper, we review some important contributions and highlight recent work.

To motivate the methods, we consider data from an environmental study which analyzed levels of perfluoroalkyl acids (PFAAs) in the cord blood of newborns born in Ottawa, Canada (Arbuckle *et al.* [7]). Exposure to PFAAs originates from both domestic and industrial products such as surfactants, fire retardants, stain-resistant coatings, and insecticides. One particular PFAA is perfluorooctanoic acid (PFOA) which has recently been confirmed to pass through the placenta. However, scientists have noted levels of certain contaminants are thought to decrease as the number of pregnancies (gravida) increases, and have virtually no effect after a certain number of pregnancies. Thus, the effect of gravida on PFOA levels displays a natural ordering, and constrained statistical inference is appropriate.

The early achievements of Brunk [8] considered maximum likelihood estimation, while Bartholomew [9–12] developed a likelihood ratio test for equality of several normal means against ordered alternatives. Kudo [13], Dykstra [14], El Barmi and Dykstra [15,16], El Barmi and Johnson [17] and Dardanoni and Forcina [18] focused on inferences under the normal or multinomial setting. Gourieroux *et al.* [19] and Shapiro [20] considered estimation and testing under linear inequality restrictions in Gaussian linear models. In the context of constrained inference for generalized linear models, important papers by Piergosch [21], Silvapulle [22] and Fahrmeir and Kligner [23] detailed the properties of a one-sided or linear inequality hypothesis test of the regression parameters. The asymptotic null distribution for the ordered hypothesis was found to be chi-bar-squared, which is a mixture of chi-squared dis-

tributions. Recent extentions ro non-normal models are discussed in later sections. Common notation used in constrained inference is presented in the following section.

Constraints of Interest

Using the terminology of statistical hypothesis testing, Molenberghs and Verbeke [2] note the two tests

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} > \mathbf{0} \quad \text{and} \tag{1}$$

$$H_0 : \boldsymbol{\theta} \leq \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} > \mathbf{0} \tag{2}$$

are distinct. In constrained inference, interest lies in (1) which does not permit any negative values in the estimate. If such a negative value occurs, it is replaced by the boundary value $\hat{\boldsymbol{\theta}} = \mathbf{0}$. Test statistics, null distributions and thus p-values differ from standard unconstrained statistical techniques.

More generally, inequality constrained testing problems may be organized into three hypotheses:

$$H_0 : \boldsymbol{\theta} \in \mathcal{M}; \quad H_1 : \boldsymbol{\theta} \in \mathcal{C}; \quad H_2 : \boldsymbol{\theta} \notin \mathcal{C} \tag{3}$$

where $\mathcal{M}$ is a linear space, $\mathcal{C}$ is a closed convex cone in the Euclidean space and $\mathcal{M} \in \mathcal{C}$. Many familiar statistical tests such as $H_0 : A\boldsymbol{\beta} = \mathbf{0}$ and $H_0 : \mu_1 = \cdots = \mu_k$ are included in $\mathcal{M}$. The definition (3) permits combinations of one- and two-sided tests to be included under a general framework.

This paper is organized as follows: Section 2 describes estimation and hypothesis testing methods for constrained inference with one-way normal models, multivariate normal models as well as non-normal models. Section 3 describes an application of constrained inference to the Ottawa cord blood study mentioned previously. Finally, Section 4 provides suggestions for future research directions.

2. ESTIMATION AND HYPOTHESIS TESTING WITH CONSTRAINTS

2.1. Univariate Normal Mean Models

Denote a set of increasing dose levels by 1, 2, ..., k where 1 corresponds to the zero or control dose level. A one-way Analysis of Variance (ANOVA) model considers n_i experimental units tested at the ith dose level, $i = 1, \ldots, k$. Let observations y_{ij} be mutually independent with $y_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. Then $\bar{y}_i \sim N(\mu_i, \sigma^2/n_i)$, $i = 1, \ldots, k$ are the sample means of the dose groups and let $S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2/\nu$ be an unbiased estimate of the common variance σ^2, with $\nu = \sum_{i=1}^{k} n_i - k > 0$. Then S^2 is distributed as $\sigma^2 \chi^2_\nu/\nu$, independently of $\bar{y}_1, \ldots, \bar{y}_k$. The parameter space for this problem is defined as $\Omega = \{\boldsymbol{\mu} \in R^k : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k\}$, with σ^2 as a nuisance parameter. The space Ω is known as the *simple order* in the constrained literature, since it denotes a non-decreasing tendency among group means.

The restricted maximum likelihood estimator of $\boldsymbol{\mu}$ subject to Ω is denoted by $\boldsymbol{\mu}^\star = (\mu_1^\star, \ldots, \mu_k^\star)$ and is defined as the isotonic regression of $\bar{\mathbf{y}} = (\bar{y}_1, \ldots, \bar{y}_k)$ under Ω with sample sizes n_1, ..., n_k. As the observations are assumed to be normally distributed, the maximum likelihood estimate (MLE) is the solution to the following

constrained weighted least squares problem:

$$\min_{\mu\in\Omega}\sum_{i=1}^{k} n_i(\bar{y}_i-\mu_i)^2. \tag{4}$$

The MLE is readily calculated using the Pool-Adjacent-Violators Algorithm (PAVA) (see Robertson, Wright and Dykstra [5]). The process is essentially a successive averaging of $\bar{y}_i$'s until a sequence of non–decreasing values is obtained, and the MLE is represented as

$$\mu_j^\star=\max_{i\le j}\min_{l\ge j}A(i,l),\quad j=1,\ldots,k\quad\text{where}\quad A(i,l)=\frac{\sum_{m=i}^{l}n_m y_m}{\sum_{m=i}^{l}n_m}.$$

The MLE of μ may then be partitioned into consecutive sequences of equal–valued $\mu_j^\star$'s such that

$$\mu_1^\star=\cdots=\mu_{i_1}^\star<\mu_{i_1+1}^\star=\cdots=\mu_{i_2}^\star<\cdots<\mu_{i_{l-1}+1}^\star=\cdots=\mu_k^\star. \tag{5}$$

As is often the case in applications, a researcher may believe that the response means are monotone increasing, *a priori*, thus likelihood ratio tests (LRTs) for homogeneity of normal means with simple order restrictions are introduced. Under the monotonicity assumption $\mu_1\le\cdots\le\mu_k$, the LRT for ordered alternatives considers the hypotheses:

$$H_0:\mu_1=\mu_2=\cdots=\mu_k;\ H_1:\mu_1\le\mu_2\le\cdots\le\mu_k;\ H_2:\text{No restrictions on }\mu_i\text{'s}. \tag{6}$$

The LRT rejects H_0 in favour of H_1-H_0 for large values of the test statistic

$$S_{01}\ =\ \frac{\displaystyle\sum_{i=1}^{k}n_i(\mu_i^\star-\bar{\mu})^2}{\displaystyle\sum_{i=1}^{k}n_i(\bar{y}_i-\mu_i^\star)^2/\nu+S^2},$$

where $\bar{\mu}=\sum_{i=1}^{k}n_i\bar{y}_i/\sum_{i=1}^{k}n_i$, the overall sample mean. When σ^2 is known, the test statistic is given by

$$\bar{\chi}_{01}^2\ =\ \frac{\displaystyle\sum_{i=1}^{k}n_i(\mu_i^\star-\bar{\mu})^2}{\sigma^2}. \tag{7}$$

As shown in [5], as $\nu\to\infty$, the distribution of S_{01} approaches that of $\bar{\chi}_{01}^2$. Similarly, test statistics for testing H_1 against H_2-H_1 are given by

$$S_{12}\ =\ \frac{\displaystyle\sum_{i=1}^{k}n_i(\bar{y}_i-\mu_i^\star)^2}{S^2},\quad\text{and}\quad\bar{\chi}_{12}^2\ =\ \frac{\displaystyle\sum_{i=1}^{k}n_i(\bar{y}_i-\mu_i^\star)^2}{\sigma^2}. \tag{8}$$

Alternatively, when σ^2 is unknown, the LRT for testing H_0 against H_1-H_0 and H_1 against H_2-H_1 are, respectively

$$\bar{E}_{01}^2=\frac{\bar{\chi}_{12}^2}{\bar{\chi}_{01}^2+\bar{\chi}_{12}^2+\nu S^2}\quad\text{and}\quad\bar{E}_{12}^2=\frac{\bar{\chi}_{12}^2}{\bar{\chi}_{12}^2+\nu S^2}. \tag{9}$$

The null distributions of $\bar{\chi}^2_{01}$, $\bar{\chi}^2_{12}$, $\bar{E}^2_{01}$, and $\bar{E}^2_{12}$ are

$$\begin{aligned}
P[\bar{\chi}^2_{01} \geq s] &= \sum_{j=1}^{k} P_s(j,k;\mathbf{n}) P[\chi^2_{j-1} \geq s] \\
P[\bar{\chi}^2_{12} \geq s] &= \sum_{j=1}^{k} P_s(j,k;\mathbf{n}) P[\chi^2_{k-j} \geq s] \\
P[\bar{E}_{01} \geq s] &= \sum_{j=1}^{k} P_s(j,k;\mathbf{n}) P[B_{(j-1)/2,(N-j)/2} \geq s] \\
P[\bar{E}_{12} \geq s] &= \sum_{j=1}^{k} P_s(j,k;\mathbf{n}) P[B_{(k-j)/2,(N-k)/2} \geq s]
\end{aligned}$$

for any $s > 0$, where $N = \sum_{i=1}^{k} n_i$, $\mathbf{n} = (n_1, \ldots, n_k)$, $P_s(j,k;\mathbf{n})$ is the level probability under H_0 that $\boldsymbol{\mu}^\star$ takes j distinct values, $B_{a,b}$ is a Beta-distribution with parameters a and b, and χ^2_c is a chi–squared variable with c degrees of freedom. For the case of equal weights, the level probabilities and the critical values are tabled in [5]. The null distributions of $\bar{\chi}^2_{01}$ and $\bar{\chi}^2_{12}$ (and $\bar{E}^2_{01}$, and $\bar{E}^2_{12}$) are essentially weighted averages of chi-squared (beta) distributions. Hence, $\bar{\chi}^2_{01}$ and $\bar{\chi}^2_{12}$ are denoted the *chi-bar-square* distributions, while $\bar{E}^2_{01}$ and $\bar{E}^2_{12}$ are denoted *E-bar-square* distributions. Both distributions play a prominent role in constrained inference.

For the simply ordered case, i.e. with $\mu_1 \leq \cdots \leq \mu_k$, the level probabilities are denoted $P_s(l,k;\mathbf{n})$. When $k = 2$, the level probabilities are $P_s(1,2;\mathbf{n}) = P_s(2,2;\mathbf{n}) = \frac{1}{2}$. In general, no closed form of level probabilities are available for arbitrary sample sizes, however, if the sample sizes are equal, the level probabilities are more readily obtained (see [5]).

Similar results are available for testing problems with other orderings. In those cases, level probabilities are completely unknown unless k is very small even if the weights are equal. Other analyses such as multiple comparisons under order restrictions are also possible. More detail on constrained inference for the one-way analysis of variance model is provided in the excellent review paper by Dobler [24].

2.2. Multivariate Normal Mean Models

Standard statistical techniques such as linear regression analysis extend the results of the one-way model and involve a response vector $\boldsymbol{y}$, which is assumed to be multivariate normal with mean vector $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{V}$. Constraints may be placed on either $\boldsymbol{\theta}$ or $\boldsymbol{V}$. If we let $\mathcal{C}$ be a closed convex cone of R^p, it is well-known that the constrained maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ under $\mathcal{C}$ (denoted $\boldsymbol{\theta}^*$) is the least squares projection of $\boldsymbol{y}$, i.e. $\boldsymbol{\theta}^* = P_{\boldsymbol{V}}(\boldsymbol{y}|\mathcal{C}) = \min_{\boldsymbol{\theta}\in\mathcal{C}} (\boldsymbol{y}-\boldsymbol{\theta})^T \boldsymbol{V}^{-1} (\boldsymbol{y}-\boldsymbol{\theta})$.

In many cases, determination of $\boldsymbol{\theta}^*$ is performed using quadratic programming. However, some algorithms have been proposed in the literature based on the least square method. Kudo [13] proposed a general algorithm which under certain closed convex cones may be considered a generalized PAVA algorithm. Dykstra [14] proposed an iterative algorithm to obtain the projection onto a closed convex cone, which is computationally faster. El Barmi and Dykstra [15] further considered an algorithm for constrained estimation of multinomial parameters using the Fenchel duality.

For hypothesis testing under constraints, consider the special case of (3) in which $H_0 : \boldsymbol{\theta} \in \mathcal{M}_0 = \{\mathbf{z} \in R^p : \mathbf{z} = \mathbf{0}\}$ versus $H_1 - H_0$ where $H_1 : \boldsymbol{\theta} \in \mathcal{C}$. Then, with $\boldsymbol{V}$ known, the likelihood ratio test (LRT) is given by

$$\bar{\chi}^2_{01}(\boldsymbol{V}, \mathcal{C}) = \mathbf{y}^{\mathrm{T}}\boldsymbol{V}^{-1}\mathbf{y} - \min_{\boldsymbol{\theta}\in\mathcal{C}}(\mathbf{y}-\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{V}^{-1}(\mathbf{y}-\boldsymbol{\theta}) = ||P(\mathbf{y}|\mathcal{C})||^2_{\boldsymbol{V}} \tag{10}$$

where the projection is taken with respect to the matrix $\boldsymbol{V}^{-1}$. When testing $H_1 : \boldsymbol{\theta} \in \mathcal{C}$ against $H_2 - H_1$, where H_2 imposes no restriction on $\boldsymbol{\theta}$, the test statistic is

$$\bar{\chi}^2_{12}(\boldsymbol{V}, \mathcal{C}) = \min_{\boldsymbol{\theta}\in\mathcal{C}}(\mathbf{y}-\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{V}^{-1}(\mathbf{y}-\boldsymbol{\theta}) = ||\mathbf{y} - P(\mathbf{y}|\mathcal{C})||^2_{\boldsymbol{V}}. \tag{11}$$

Hence, the chi-bar-square test statistics are expressed in terms of the distance between the origin of $\mathbf{y}$ and its projection onto a closed convex cone. Then, with $\boldsymbol{V}$ a $p \times p$ positive definite matrix, under H_0 we have

$$pr\{\bar{\chi}^2_{01}(\boldsymbol{V}, \mathcal{C}) \leq c\} = \sum_{i=0}^{p} w_i(p, \boldsymbol{V}, \mathcal{C}) pr(\chi^2_i \leq c), \tag{12}$$

$$pr\{\bar{\chi}^2_{12}(\boldsymbol{V}, \mathcal{C}) \leq c\} = \sum_{i=0}^{p} w_{p-i}(p, \boldsymbol{V}, \mathcal{C}) pr(\chi^2_i \leq c), \tag{13}$$

where $w_i(p, \boldsymbol{V}, \mathcal{C})$, $i = 0, \ldots, p$ are some nonnegative numbers and $\sum_{i=0}^{p} w_i(p, \boldsymbol{V}, \mathcal{C}) = 1$. The right hand side of equation (12) is the *chi-bar-square distribution*, and is a weighted mean of several tail probabilities of χ^2 distributions. The set $\{w_i(p, \boldsymbol{V}, \mathcal{C})\}$ is known as the *chi-bar-square weights* or simply *weights*. Similar results may be derived when the null space is replace by a linear space constrained in $\mathcal{C}$, such as $A\boldsymbol{\theta} = \mathbf{c}$, where A is a $q \times p$ matrix and $\mathbf{c}$ is a $q \times 1$ vector.

Interestingly, a distinguishing factor of the $\bar{\chi}^2_{12}$ test is that the null hypothesis involves inequalities. Hence, the p-value depends on the underlying parameter $\boldsymbol{\theta}$, which may be anywhere in the null parameter space. As an example, with $\mathcal{C} = \{\boldsymbol{\theta} : \boldsymbol{R}_1\boldsymbol{\theta} \geq \mathbf{0}\}$, in order to obtain the critical value, c which assures size α, we must solve $\sup_{\boldsymbol{R}_1\boldsymbol{\theta}\geq\mathbf{0}} P_{\boldsymbol{\theta}}[\bar{\chi}^2_{12} > c] = \alpha$, where $\boldsymbol{R}_1$ is a $q_1 \times p$ matrix. As explained in Silvapulle and Sen [6], the supremum occurs at any $\boldsymbol{\theta}_0$ with $\boldsymbol{R}_1\boldsymbol{\theta}_0 = \mathbf{0}$, and hence $\boldsymbol{\theta} = \mathbf{0}$ is one such value. This particular null distribution is denoted the least favorable distribution.

Furthermore, analogous to the level probabilities discussed previously, closed form expressions for w_i exist only when the number of parameters is small (i.e. $p \leq 4$) (Kudo [13], Shapiro [20]). If $p \geq 5$, simulated weights may be used. Approaches to simulate the chi-bar-square weights are provided in [6] (pp. 78-80).

Moreover, as $\boldsymbol{V}$ is often unknown in practice, we assume the covariance matrix has the form $V = \sigma^2\boldsymbol{U}$, with $\boldsymbol{U}$ known and σ^2 unknown. Suppose we consider the linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{E}$ where $\boldsymbol{X}$ is a known $N \times p$ matrix of rank p and $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, with $\boldsymbol{E} \sim N(\mathbf{0}, \sigma^2\boldsymbol{U})$.We then have

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{U}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{U}^{-1}\boldsymbol{Y}, \quad \text{and} \quad Q(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta})^T\boldsymbol{U}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta}),$$

where $\hat{\boldsymbol{\theta}}$ represents the unconstrained estimator of $\boldsymbol{\theta}$. Similar to the results in the previous section, if we define the hypotheses $H_a : \boldsymbol{\theta} \in \mathcal{C}_a$ and $H_b : \boldsymbol{\theta} \in \mathcal{C}_b$ with $\mathcal{C}_a \subset \mathcal{C}_b$, then $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ are the values over which $Q(\boldsymbol{\theta})$ is minimized over $\mathcal{C}_a$ and $\mathcal{C}_b$ respectively. Then, the E-bar-square statistic is

$$\bar{E}^2_{ab} = \{Q(\boldsymbol{\theta}_a) - Q(\boldsymbol{\theta}_b)\}/Q(\boldsymbol{\theta}_a). \tag{14}$$

Under the assumption of normality, $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ are the maximum likelihood estimates of $\boldsymbol{\theta}$ under H_a and H_b respectively. Furthermore it may be shown that $\bar{E}^2$ is equivalent to the likelihood ratio test when the error vector $\boldsymbol{E} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{U})$.

The null distributions of $\bar{E}^2_{01}$ and $\bar{E}^2_{12}$ are formed similarly to the corresponding chi-bar square distributions. For $H_0 : \boldsymbol{\theta} \in \mathcal{M}$ versus $H_1 - H_0$ where $H_1 : \boldsymbol{\theta} \in \mathcal{C}$, and $dim(\mathcal{M}) = q$, then the least favourable null value for $\bar{E}^2_{12}$ is $\boldsymbol{\theta} = \mathbf{0}$, as before. The null distributions are

$$\begin{aligned} pr\{\bar{E}^2_{01} \leq c\} &= \sum_{i=0}^{p} w_i(p, (\boldsymbol{X}^T \boldsymbol{U}^{-1} \boldsymbol{X})^{-1}, \mathcal{C}) pr[B\{i/2, (N-q-i)/2\} \leq c] \\ pr\{\bar{E}^2_{12} \leq c | \boldsymbol{\theta} = \mathbf{0}\} &= \sum_{i=0}^{p} w_{p-i}(p, (\boldsymbol{X}^T \boldsymbol{U}^{-1} \boldsymbol{X})^{-1}, \mathcal{C}) pr[B\{i/2, (N-p)/2\} \leq c], \end{aligned}$$

where $B[a, b]$ is a beta distribution with parameters a and b.

Note that when $\boldsymbol{V}$ is completely unknown, Perlman [25] derived the exact likelihood ratio test. Extensions to this test have been considered, and are summarized in [6].

2.3. Recent Extensions to Constrained Inference

Development of constrained methods for non-normal models such as generalized linear models (logistic and Poisson regression, proportional hazard models), time series models, etc. have been undertaken by many authors (Piergosh [21], Silvapulle [22], and Fahrmeir & Kligner [23]) as mentioned previously. Under additional regularity conditions and by implementing the inverse of the Fisher information matrix, the null distributions of the LRTs for the hypotheses given in (3) also follow a chi-bar-square distribution, albeit asymptotically for these models. Moreover, as in the unconstrained case, constrained versions of the LRT, score and Wald tests are also available and may be shown to be asymptotically equivalent [2,6]. Hall and Praestgaard [26] introduced an order-restricted score test for homogeneity in the generalized linear mixed model, which accounts for the positive semidefinite assumption of the random effects covariance matrix. Park *et al.* [27], Pilla *et al.* [28] and Rosen and Davidov [29] considered constrained estimation of parameters in a longitudinal setup for which the chi-bar-square distributions also apply. Also, Farrell and Park [30] proposed a constrained likelihood ratio test for ordered group effects with one binary and one continuous response, which was shown to follow a chi-bar-square distribution asymptotically.

In the context of linear modeling, estimation of order-restricted model parameters with incomplete data has been studied previously (Kim & Taylor [31]; Shi, Zheng & Guo [32]; Zheng, Shi & Guo [33]). For linear mixture models, Jamshidian [34] used a somewhat different approach, and proposed a globally convergent algorithm based on the gradient projection (GP) method which may be employed as part of the Expectation -Maximization (EM) algorithm (Dempster *et al.*, [35]) for estimation under incomplete data. However, no distributional results were provided. Nettleton [36] discusses various theoretical issues relating to the convergence properties of the EM procedure under order-restrictions. The aforementioned papers also conducted simulation studies which demonstrated constrained techniques tended to have smaller mean square error than unrestricted models which ignored the constraints.

With the recent advancement of mixed models and missing-data analysis, interest lies in obtaining constrained estimates and test statistics in such situations. However, unlike generalized linear models in which many models may be approximated by a quadratic function and multivariate normal techniques used; a generalized linear mixed model (GLMM) setting is known to lose efficiency when such approximations are performed. Hence, for constrained estimation of mixed models and for missing data problems, a different approach must be considered for estimation, since constrained algorithms based on a least square method such as Dykstra [14] may not be applied. Recent papers by Jamshidian [34] and Davis *et al.* [37] have implemented modified versions of the gradient projection algorithm to find maximum likelihood estimates for non-normal models under inequality constraints.

The gradient projection algorithm is known in the optimization literature as an active set method for solving equations under linear inequality constraints. As stated by Luenberger [38], if it were known *a priori* which constraints were active (hold with equality) at the solution to the optimization problem, the solution would be a local maximum point of the problem defined by ignoring the inactive constraints and treating all active constraints as equality constraints. Hence, with respect to local or relative solutions, the problem could be regarded as having equality constraints only. This observation suggests that the majority of theory applicable to the optimization problem may be derived by considering the equality constraints alone. Such estimates then satisfy the Kuhn-Tucker [39] first-order necessary conditions for optimality.

With $\mathcal{M} = \{\boldsymbol{\beta} : A\boldsymbol{\theta} = \mathbf{c}\}$ and $\mathcal{C} = \{\boldsymbol{\theta} : A\boldsymbol{\theta} \leq \mathbf{c}\}$, likelihood ratio test statistics may be calculated using constrained estimates under H_0 and H_1, denoted $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^*$ respectively. The unconstrained test rejects H_0 in favor of $H_2 - H_0$ for large values of the test statistic $T_{02} = 2[l(\hat{\boldsymbol{\theta}}|\mathbf{y}) - l(\boldsymbol{\theta}^0|\mathbf{y})]$, where $l(\boldsymbol{\theta}|\mathbf{y})$ is the logarithm of the marginal likelihood. It is well-known that under H_0 the statistic T_{02} asymptotically follows a chi-square distribution with q degrees of freedom, assuming A is a $q \times p$ matrix. However, if the parameter space is restricted by H_1, we test H_0 against $H_1 - H_0$ using the statistic $T_{01} = 2[l(\boldsymbol{\theta}^*|\mathbf{y}) - l(\boldsymbol{\theta}^0|\mathbf{y})]$. Further, we first confirm H_1 by a goodness-of-fit test which rejects H_1 for large values of $T_{12} = 2[l(\hat{\boldsymbol{\theta}}|\mathbf{y}) - l(\boldsymbol{\theta}^*|\mathbf{y})]$.

Davis *et al.* [37] showed that additional theoretical rigour is needed to prove that the asymptotic null distributions of the likelihood ratio test statistics T_{01} and T_{12} for GLMMs with cluster correlated data are:

$$\lim_{k\to\infty} P_{\boldsymbol{\theta}_0}[T_{01} > x] = \sum_{i=0}^{q} w_i(q, AV(\boldsymbol{\theta}_0)A^T)P[\chi_i^2 > x], \tag{15}$$

and

$$\lim_{k\to\infty} P_{\boldsymbol{\theta}_0}[T_{12} > x] = \sum_{i=0}^{q} w_{q-i}(q, AV(\boldsymbol{\theta}_0)A^T)P[\chi_i^2 > x], \tag{16}$$

for any $x \geq 0$. Here q is the rank of A, k is the number of clusters, $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$ is a value of $\boldsymbol{\theta}$ under H_0, $\boldsymbol{\gamma}_0$ represents the variance components, and $V(\boldsymbol{\theta}_0) = \lim_{k\to\infty} k[\mathcal{I}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) - \mathcal{I}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\mathcal{I}^{-1}(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0)]^{-1}$, where $\mathcal{I}(.,.) = E[\mathcal{I}_o(.,.)]$ with $\mathcal{I}_o(.,.)$ being the observed information matrix. The chi-bar-square weights, $w_i(q, D)$, represent the probability that the least squares projection of a q-dimensional multivariate normal observation from $N(\mathbf{0}, D)$ onto the positive or-

thant cone has exactly i positive component values. As in the multivariate normal case, simulation algorithms may be used to find $w_i(q, D)$.

3. APPLICATION: OTTAWA CORD BLOOD STUDY

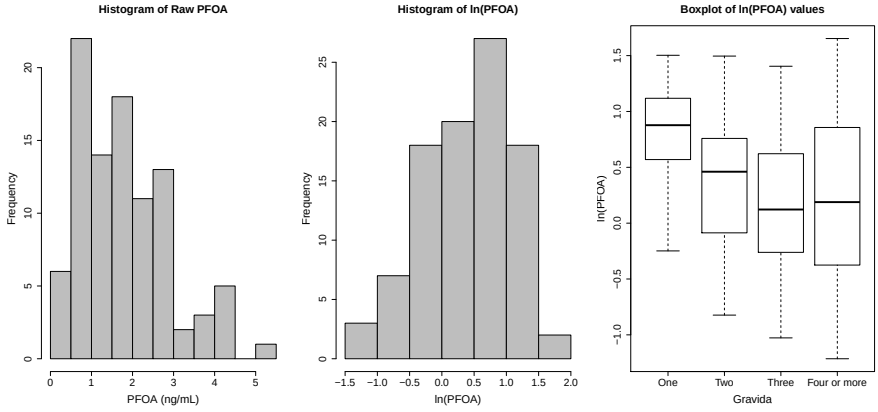

Figure 1
Histograms and Boxplot of PFOA Levels in Ottawa Cord Blood Study

We revisit the environmental study mentioned in the Introduction, which considered levels of PFOA in the umbilical cord blood of newborn babies. A total of 126 mothers participated in the study, of which 100 of the cord serums were of sufficient volume for analysis. Details of the study are provided in Arbuckle *et al.* [7]. We extend the results of the previous analysis by incorporating constrained statistical techniques. Table 1 provides descriptive statistics of PFOA levels (ng/mL) by gravida for the cord blood samples. As is typical in environmental studies, the lognormal distribution is used to account for skewness (Figure 1). If Y is a lognormal random variable, then the natural logarithm of Y is a normal random variable. Table 1 and the boxplot in Figure 1 note the decreasing relationship between gravida and PFOA levels.

Table 1
Descriptive Statistics for PFOA (ng/mL) by Gravida for Ottawa Cord Blood Study

Gravida	n	Arithmetic mean	SD	Min	Max	Q1	Median	Q3	Geometric mean	95%CI for geometric mean
One	23	2.56	1.08	0.78	4.50	1.61	2.40	3.10	2.31	(1.90, 2.82)
Two	33	1.70	0.92	0.44	4.46	0.93	1.59	2.15	1.47	(1.22, 1.78)
Three	18	1.43	0.93	0.36	4.08	0.77	1.12	1.86	1.19	(0.90, 1.59)
Four or more	26	1.47	1.13	0.30	5.22	0.62	1.10	2.18	1.13	(0.85, 1.51)

The statistical model is

$$\ln y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \epsilon_i$$

where x_{i1}, x_{i2}, x_{i3} are indicator variables representing gravida of one, two and three, respectively; x_{i4} had a value of 1 if the delivery is Vaginal or 0 if by Cesarean section; x_{i5} has a value of 1 if the birth weight is less than 2500g and 0 otherwise; and assume $\epsilon_i \sim N(0, \sigma^2)$. Values of the unconstrained parameter estimates are provided in Table 2. The statistical package R was used for computation.

In order to test for a decreasing gravida effect and no difference in the effect of the third and subsequent pregnancy, we define the parameter spaces $\mathcal{M}_0 = \{\boldsymbol{\beta} : \beta_1 = \beta_2 = \beta_3 = 0\}$ under H_0 and $\mathcal{C}_1 = \{\boldsymbol{\beta} : \beta_1 \geq \beta_2 \geq \beta_3 = 0\}$ under H_1. The unconstrained parameter space under H_2 is $\mathcal{C}_2 = \{\boldsymbol{\beta} : \boldsymbol{\beta} \in R^p\}$. Since the unconstrained estimates do not satisfy either constrained parameter space, the gradient projection algorithm is used to fit maximum likelihood estimates under $\mathcal{M}_0$ and $\mathcal{C}_1$, as listed in Table 2. A description of the algorithm is provided in the Appendix.

To perform hypothesis tests, the unconstrained test statistic was found to be $\bar{E}^2_{02} = 0.09246$ with a p-value of 0.0002. The goodness-of-fit test of H_1 versus $H_2 - H_1$ had a test statistic $\bar{E}^2_{12} = 0.00001$ with p-value = 0.983, thus the constrained hypothesis test is useful. The test of H_0 versus $H_1 - H_0$ has test statistic $\bar{E}^2_{01} = 0.09245$ with p-value = 0.00004. Hence, gravida has a decreasing effect on PFOA levels and the third and subsequent pregnancy has no significant effect. Thus both constrained and unconstrained tests lead to a rejection of the null hypothesis, however the additional information provided by the constrained test leads to a more efficient and lower p-value, as well as a more precise interpretation.

The aforementioned p-values were obtained using E-bar-square distributions with associated weights of $(0.32730, 0.49522, 0.17748)^T$, which were obtained by simulation. Note that to avoid redundancy, the weight vector has been reduced by one dimension in this case, since there is an equality constraint in $\mathcal{C}_1$.

Table 2
Constrained and Unconstrained ML Estimates for Ottawa Cord Blood Study

Parameter	Unconstrained		Constrained	
	$\mathcal{C}_2$ $\{\boldsymbol{\beta} : \boldsymbol{\beta} \in \boldsymbol{R^p}\}$	Standard error[a]	$\mathcal{M}_0$ $\{\boldsymbol{\beta} : \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \boldsymbol{\beta_3} = \mathbf{0}\}$[b]	$\mathcal{C}_1$ $\{\boldsymbol{\beta} : \boldsymbol{\beta_1} \geq \boldsymbol{\beta_2} \geq \boldsymbol{\beta_3} = \mathbf{0}\}$[b]
β_0	0.121	0.121	0.254	0.119
β_1	0.485	0.181	0.000	0.486
β_2	0.160	0.160	0.000	0.162
β_3	-0.003	0.186	0.000	0.000
β_4	0.444	0.152	0.595	0.444
β_5	0.469	0.344	0.501	0.470
σ^2	0.320	0.046	0.354	0.321

[a] where the standard error refers to the unconstrained value.

[b] estimates obtained using the Gradient Projection algorithm.

4. FUTURE DIRECTIONS

While constrained statistical inference requires additional computational and theoretical rigour, the improvements in efficiency and additional testing power are well-established. Nevertheless, many unsolved problems remain. Contrary to unconstrained models, hypothesis testing under the chi-bar-square or E-bar-square distribution is the preferred method of inference as confidence intervals are difficult to obtain in constrained environments. The distribution of the constrained estimator $\tilde{\boldsymbol{\theta}}$ is dependent upon the proximity of the unconstrained estimator to the boundary of the constraint, and has been derived for linear models under simple orderings, as per Hwang and Peddada [40]. Additional theoretical advances would be needed to obtain confidence intervals for more complicated models in constrained settings.

While much research in constrained inference has focused on frequentist approaches, Dunson and Neelon [41] considered Bayesian constrained methods for GLMs, highlighting the usefulness of Bayesian constrained procedures. With continual advances in computational methods, Bayesian constrained techniques are a timely and useful area of future research. The authors noted that sampling from the constrained posterior distribution is obtained by transforming draws from the unconstrained posterior density. As a result, existing Gibbs sampling algorithms for posterior computation of generalized linear models apply directly. Bayesian inferences for umbrella orderings were discussed in Hans and Dunson [42].

Other constraints such as shape or stochastic orders may also be considered, as discussed by Silvapulle and Sen [6], Peddada *et al.* [43] Bornkamp *et al.* [44] and Lee *et al.* [45]. Alvo [46] considers nonparametric tests for umbrella orderings, for which the dosage mean values increase and then decrease after a certain peak. Other approaches to maximum likelihood which require fewer assumptions, are also relevant in constrained settings, and could be extended to nonparametric methods for other parameter orderings.

Another area of future work would be to consider constraints on variance-covariance parameters. Calvin and Dykstra [47] developed a restricted maximum likelihood (REML) estimation scheme for covariance matrices, with both balanced and unbalanced data. Such an extension would be particularly useful in mixed models, for which tests for increasing or decreasing trends in variance components could be developed. Further, for missing data models, Nettleton [36] outlined a theorem for constrained estimators which extends the EM algorithm under constraints. These developments would prove particularly useful, since even in the unconstrained case, the EM algorithm is known to converge slowly compared to other maximization techniques.

ACKNOWLEDGMENTS

The author would like to thank the editor and a referee for their insightful comments which lead to an improved manuscript. The author would also like to thank Dr. Tye Arbuckle of the Population Studies Division, Health Canada, for technical assistance with respect to the Ottawa Cord blood study.

REFERENCES

[1] Peng, J., Lee, C.I.C., Davis, K.A., & Wang, W. (2008). Stepwise confidence intervals for monotone dose–response studies. *Biometrics, 64*, 877–885.

[2] Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and wald tests in a constrained parameter space. *The American Statistician, 61*, 22–27.

[3] Dawson, D., & Magee, L. (2001). The national hockey league entry draft, 1969–1995: an application of a weighted pool-adjacent violators algorithm. *The American Statistician, 55*(3), 194–199.

[4] Barlow, R.E., Bartholomew, D.J., Bremner, J.M., & Brunk, H.D. (1972). *Statistical inference under order restrictions: the theory and application of isotonic regression*. New York: Wiley.

[5] Robertson, T., Wright, F.T., & Dykstra, R.L. (1988). *Order restricted statistical inference*. Chichester: John Wiley.

[6] Silvapulle, M.J., & Sen, P.K. (2005). *Constrained statistical inference.* Hoboken: John Wiley.

[7] Arbuckle, T.E., Kubwabo, C., Walker, M., Davis, K., Lalonde, K., Kosarac, I., Wen, S.W., & Arnold, D.L. (2012). Umbilical cord blood levels of perfluoroalkyl acids and polybrominated flame retardants. *International Journal of Hygiene and Environmental Health*. doi: 10.106/j.ijeh.2012.03.004

[8] Brunk, H.D. (1955). Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics, 26*, 606–616.

[9] Bartholomew, D.J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika, 46*, 36–48.

[10] Bartholomew, D.J. (1959b). A test of homogeneity for ordered alternatives II. *Biometrika, 46*, 328–335.

[11] Bartholomew, D.J. (1961a). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society, Series B, 23*(2), 239–281.

[12] Bartholomew, D.J. (1961b). Ordered tests in the analysis of variance. *Biometrika, 48*(3/4), 325–332.

[13] Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika, 50*, 403–418.

[14] Dykstra, R. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association, 78*, 837–842.

[15] El Barmi, H., & Dykstra, R. (1994). Restricted multinomial maximum likelihood estimation based upon fenchel duality. *Statistics and Probability Letters, 21*, 121–130.

[16] El Barmi, H., & Dykstra, R. (1995). Testing for and against a set of linear inequality constraints in a multinomial setting. *The Canadian Journal of Statistics, 23*, 131–143.

[17] El Barmi, H., & Johnson, M. (2006). A unified approach to testing for and against a set of linear inequality constraints in the product multinomial setting. *Journal of Multivariate Analysis, 97*, 1894–1912.

[18] Dardanoni, V., & Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association, 93*, 1112–1123.

[19] Gourieroux, C., Holly, A., & Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker-test in linear models with inequality constraints on the regression parameters. *Econometrica, 50*, 63–80.

[20] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review, 56*, 49–62.
[21] Piergosch, W.W. (1990). One-sided significance tests for generalized linear models under dichotomous response. *Biometrics, 46*, 309–316.
[22] Silvapulle, M.J. (1994). On tests against one-sided hypotheses in some generalized linear models. *Biometrics, 50*, 853–858.
[23] Fahrmeir, L., & Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers, 35*, 211–229.
[24] Dobler, C.P. (2002). The one-way layout with ordered parameters: a survey of advances since 1988. *Journal of Statistical Planning and Inference, 107*, 75–88.
[25] Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics, 40*, 549–567.
[26] Hall, D.B., & Praestgaard, J.T. (2001). Order-restricted score tests for homogeneity in generalised linear and nonlinear mixed models. *Biometrika, 88*, 739–751.
[27] Park, T., Shin, D.W., & Park, C.G. (1998). A generalized estimating equations approach for testing ordered group effects with repeated measurements. *Biometrics, 54*, 1645–1653.
[28] Pilla, R.S., Qu, A., & Loader, C. (2006). Testing for order–restricted hypotheses in longitudinal data. *Journal of the Royal Statistical Society - Series B, 68*, 437–455.
[29] Rosen, S., & Davidov, O. (2012). Order-restricted inference for multivariate longitudinal data with applications to the natural history of hearing loss. *Statistics in Medicine*. doi: 10.1002/sim.5335.
[30] Farrell, P.J., & Park, C.G. (2007). Testing for ordered group effects in binary and continuous outcomes. *Biometrical Journal, 49*, 585–598.
[31] Kim, D.K., & Taylor, J.M.G. (1995). The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. *Journal of the American Statistical Association, 90*, 708–716.
[32] Shi, N.Z., Zheng, S.R., & Guo, J. (2005). The restricted EM algorithm under inequality restrictions on the parameters. *Journal of Multivariate Analysis, 92*, 53–76.
[33] Zheng, S.R., Shi, N.Z., & Guo, J. (2005). The restricted EM algorithm under linear inequalities in a linear model with missing data. *Science in China Series A - Mathematics, 48*, 819–828.
[34] Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Computational Statistics & Data Analysis, 45*, 137–157.
[35] Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.
[36] Nettleton, D. (1999). Convergence properties of the EM algorithm in constrained parameter spaces. *The Canadian Journal of Statistics, 27*, 639–648.
[37] Davis, K.A., Park, C.G., & Sinha, S.K. (2012). Testing for generalized linear mixed models with cluster correlated data under linear inequality constraints. *The Canadian Journal of Statistics, 40*, 243–258.
[38] Luenberger, D.G. (2003). *Linear and nonlinear programming* (2nd ed.). New York: Springer.
[39] Kuhn, H.W., & Tucker, A.W. (1951). Nonlinear programming. *2nd Berkeley*

Symposium, Math. Statist. Probab. (pp. 481–492). Berkeley: University of California Press.

[40] Hwang, J.T.G., & Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *The Annals of Statistics, 22*, 67–93.

[41] Dunson, D.B., & Neelon, B. (2003). Bayesian inference on order–constrained parameters in generalized linear models. *Biometrics, 59*, 286–295.

[42] Hans, C., & Dunson, D.B. (2005). Bayesian inferences on umbrella orderings. *Biometrics, 61*, 1018–1026.

[43] Peddada, S.D., Dunson, D.B., & Tan, X. (2005). Estimation of order–restricted means from correlated data. *Biometrika, 92*, 703–715.

[44] Bornkamp, B., Ickstadt, K., & Dunson, D. (2010). Stochastically ordered multiple regression. *Biostatistics, 11*, 419–431.

[45] Lee, C.I.C., Liu, W., Park, C.G., & Peng, J. (2011). Inference for comparing a multinomial distribution with a known standard. *Statistical Papers, 53*, 775–788.

[46] Alvo, M. (2008). Nonparametric tests of hypotheses for umbrella orderings. *The Canadian Journal of Statistics, 36*, 143–156.

[47] Calvin, J.A., & Dykstra, R.L. (1995). REML estimation of covariance matrices with restricted parameter spaces. *Journal of the American Statistical Association, 90*, 321–329.

APPENDIX – GRADIENT PROJECTION ALGORITHM FOR INEQUALITY CONSTRAINTS

Consider inequality constraints of the form $A\boldsymbol{\theta} \leq \mathbf{c}$, where A is a $q \times p$ matrix of full rank ($q \leq p$), thus the constrained parameter space is $\Omega = \{\boldsymbol{\theta} : A\boldsymbol{\theta} \leq \mathbf{c}\}$. Jamshidian [34] and Davis *et al.* [37] propose a gradient projection algorithm to find a solution to maximize the log-likelihood function $l(\boldsymbol{\theta}|\boldsymbol{y})$ subject to

$$\begin{aligned} a_i^{\mathrm{T}}\boldsymbol{\theta} &= c_i \quad i \in I_1, \\ a_i^{\mathrm{T}}\boldsymbol{\theta} &\leq c_i \quad i \in I_2, \end{aligned}$$

where the likelihood is assumed to be sufficiently smooth. The algorithm begins with an initial working set of active constraints, denoted $\mathcal{W}$. This set includes indexes of the constraints in I_1, if any, and may include indexes from I_2. Let $\overline{A}$ be an $\overline{m} \times p$ matrix whose rows consist of a_i^{T} for all $i \in \mathcal{W}$ and let $\overline{\mathbf{c}}$ be the corresponding vector of c_i's. We further define V as a positive-definite matrix under the current estimate of $\boldsymbol{\theta}$ (e.g. variance-covariance matrix).

Beginning with an initial point $\boldsymbol{\theta}_r$ that satisfies $\overline{A}\boldsymbol{\theta}_r = \overline{\mathbf{c}}$, the algorithm proceeds as follows:

1. Compute $\mathbf{d} = P\widetilde{s}(\boldsymbol{\theta}_r)$ where $P = I - V\overline{A}^{\mathrm{T}}(\overline{A}V\overline{A}^{\mathrm{T}})^{-1}\overline{A}$.
2. If $\mathbf{d} = \mathbf{0}$, compute the Lagrange multipliers $\boldsymbol{\lambda} = (\overline{A}V\overline{A}^{\mathrm{T}})^{-1}\overline{A}\widetilde{s}(\boldsymbol{\theta}_r)$.
   a) If $\lambda_i \geq 0$ for all $i \in \mathcal{W} \cap I_2$, stop. The current point satisfies the Kuhn-Tucker necessary conditions.
   b) If there is at least one $\lambda_i < 0$ for $i \in \mathcal{W} \cap I_2$, determine the index corresponding to the smallest such λ_i and delete the index from $\mathcal{W}$. Modify $\overline{A}$ and $\overline{\mathbf{c}}$ by dropping a row from each accordingly and go to Step 1.
3. If $\mathbf{d} \neq \mathbf{0}$, obtain $\alpha_1 = \max_\alpha\{\alpha : \boldsymbol{\theta} + \alpha\mathbf{d} \text{ is feasible}\}$. Then search for $\alpha_2 = \max_\alpha\{l(\boldsymbol{\theta} + \alpha\mathbf{d}) : 0 \leq \alpha \leq \alpha_1\}$. Set $\widetilde{\boldsymbol{\theta}_r} = \boldsymbol{\theta}_r + \alpha_2\mathbf{d}$. Add indexes of new coordinates, if any, of $\widetilde{\boldsymbol{\theta}_r}$ that are newly on the boundary to the working set $\mathcal{W}$. Modify $\overline{A}$ and $\overline{\mathbf{c}}$ by adding additional rows.
4. Replace $\boldsymbol{\theta}$ by $\widetilde{\boldsymbol{\theta}_r}$ and go to Step 1, continuing until convergence.