TRANSACTIONS OF SOCIETY OF ACTUARIES
1984 VOL. 36

A GENERALIZATION OF WHITTAKER-HENDERSON GRADUATION

FUNG YEE CHAN, LAI K. CHAN,* AND MAN HEI YU**

ABSTRACT

The Whittaker-Henderson graduation method of minimizing $F + \lambda S$ with fit F of the form $\Sigma w_x|u_x - u''_x|^p$ and smoothness S of the form $\Sigma|\Delta^z u_x|^p$ is investigated for $1 < p \leq \infty$. It is shown that for a given λ, the set of graduated values $\underset{\sim}{u}^\lambda = (u_1^\lambda, \ldots, u_n^\lambda)^T$ is unique and is the solution of the system of equations $F'(\underset{\sim}{u}) + \lambda S'(\underset{\sim}{u}) = \underset{\sim}{0}^T$ when $1 < p < \infty$, and $\underset{\sim}{u}^\lambda$ is an optimal solution of an equivalent linear programming problem when $p = \infty$. Algorithms for computing $\underset{\sim}{u}^\lambda = (u_1^\lambda, \ldots, u_n^\lambda)^T$ are proposed. The graduated values for different p are compared. Some properties of $F(\underset{\sim}{u}^\lambda)$, $S(\underset{\sim}{u}^\lambda)$ and $F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$ are obtained. The modification to the minimization of F (or S) when S (or F) is constrained to be less than or equal to a predetermined number is proposed and studied.

1. INTRODUCTION

Given a vector of ungraduated (that is, observed) values $\underset{\sim}{u}'' \equiv (u''_1, \ldots, u''_n)^T$ and a constant $\lambda \geq 0$, the Whittaker-Henderson graduation method finds the optimum values $\underset{\sim}{u}^\lambda = (u_1^\lambda, \ldots, u_n^\lambda)^T$, called the graduated values, which minimize

$$F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u}) \text{ over all } \underset{\sim}{u} \equiv (u_1, \ldots, u_n)^T,$$

where F is a measure of the fit of $\underset{\sim}{u}$ to $\underset{\sim}{u}''$ and S is a measure of the smoothness of the values in $\underset{\sim}{u}$.

The well-known Whittaker-Henderson Type B method presented in the

* Dr. Chan, not a member of the Society, is Professor and Head, Department of Statistics, University of Manitoba, Canada.

** Mr. Yu, not a member of the Society, is a graduate student in the Department of Actuarial and Management Sciences, University of Manitoba, Canada.

Society of Actuaries Part 5 Study Notes by Greville [5, pp. 49-54] uses the square of the ℓ_2-norms:

$$F(\underset{\sim}{u}) \equiv \sum_{x=1}^{n} w_x(u_x'' - u_x)^2 \text{ and } S(\underset{\sim}{u}) \equiv \sum_{x=1}^{n-z} (\Delta^z u_x)^2,$$

where the $w_x > 0$ are the weights assigned to the u_x'', *and the* $\Delta^z u_x$ are the zth differences of u_x. The formula for the graduated values is obtained elegantly by Greville [5, pp. 49-54] using linear algebra and by Shiu [15] using advanced calculus.

Schuette [14] used the ℓ_1-norms:

$$F(\underset{\sim}{u}) \equiv \sum_{x=1}^{n} w_x|u_x'' - u_x| \text{ and } S(\underset{\sim}{u}) \equiv \sum_{x=1}^{n-z} |\Delta^z u_x|$$

and showed that $\underset{\sim}{u}^{\lambda}$ can be obtained by formulating the problem as a linear programming problem.

In Section 2 of this paper, the general case of the ℓ_p-norms is solved:

$$F(\underset{\sim}{u}) \equiv \sum_{x=1}^{n} w_x|u_x'' - u_x|^p \text{ and } S(\underset{\sim}{u}) \equiv \sum_{x=1}^{n-z} |\Delta^z u_x|^p$$

with $1 < p < \infty$.

In the discussion of Schuette's paper [14], Professor Greville [6] suggested that "it would be most interesting and worthwhile if someone would perform the same task for the Chebyshev norm that Schuette has done for the ℓ_1-norm."

Before proceeding to the case $p = \infty$, we digress to discuss the definition of the ℓ_p-norm. The definition of norm is given by Schuette [14]. If $\underset{\sim}{y} = (y_1, \ldots, y_n)$ is a vector of real or complex numbers, then the ℓ_p-norm of the vector, denoted by $\|\underset{\sim}{y}\|_p$, is defined as

$$\|\underset{\sim}{y}\|_p = (|y_1|^p + \ldots + |y_n|^p)^{1/p}$$

for $1 \leq p < \infty$. For the case $p = \infty$, the ℓ_∞-norm or Chebyshev norm (also called the uniform norm) is defined as

$$\|\underset{\sim}{y}\|_\infty = \max_{1 \leq x \leq n} |y_x|.$$

It is intuitively clear and can be shown analytically that the following property holds [cf. 10, p. 248]

$$\lim_{p\to\infty} \|\underset{\sim}{y}\|_p = \|\underset{\sim}{y}\|_\infty.$$

The term $F(\underset{\sim}{u}) \equiv \sum_{x=1}^{n} w_x|u_x - u_x''|^p$ is the pth power of the weighted norm of $u_x - u_x''$ with weights w_x, and $S(\underset{\sim}{u}) \equiv \sum_{x=1}^{n-z} |\Delta^z u_x|^p$ is the pth power of the norm of $\Delta^z u_x$. Therefore, in the ℓ_∞-norm case, $F(\underset{\sim}{u})$ and $S(\underset{\sim}{u})$ should be defined as:

$$F(\underset{\sim}{u}) \equiv \max_{1\le x\le n} |u_x - u''_x| \text{ and } S(\underset{\sim}{u}) \equiv \max_{1\le x\le n-z} |\Delta^z u_x|.$$

The weights w_x disappear in the term $F(\underset{\sim}{u})$ since

$$\lim_{p\to\infty}\left(\sum_{x=1}^{n} w_x|u_x - u''_x|^p\right)^{1/p} = \max_{1\le x\le n} |u_x - u''_x|$$

which can be seen from the property indicated earlier.

In section 3 of this paper the ℓ_∞-norm problem suggested by Greville [6] is solved by formulating it as a linear programming problem. At that point, the problem of obtaining graduated values for the well-known Whittaker-Henderson graduation method will have been solved for all ℓ_p-norms with $1 \le p \le \infty$.

The ℓ_∞-norm case is further generalized to include the weights w_x, for example,

$$F(\underset{\sim}{u}) \equiv \max_{1\le x\le n} w_x|u_x - u''_x|.$$

The term $\max_{1\le x\le n} w_x|u_x - u''_x|$ is a weighted ℓ_∞-norm of $\underset{\sim}{u} - \underset{\sim}{u}''$, although it does not have the property that

$$\lim_{p\to\infty}\left(\sum_{x=1}^{n} w_x|u_x - u''_x|^p\right)^{1/p} = \max_{1\le x\le n} w_x|u_x - u''_x|.$$

Although the u_x are usually nonnegative in actuarial applications, we allow them to be negative in our studies. Minimizing a nonlinear function of several variables under constraints is theoretically and computationally complicated, because the optimal solution may occur on the boundary (for example, $u^\lambda_x = 0$ for some x). In practice, when the ungraduated values u''_x are positive, the graduated values u^λ_x will usually be positive even when the nonnegative constraints are not imposed on the u_x.

In Section 4, it is shown that the solutions have the *Monotone Properties*:

$F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$ is a nondecreasing function of λ
$F(\underset{\sim}{u}^\lambda)$ is a nondecreasing function of λ
$S(\underset{\sim}{u}^\lambda)$ is a nonincreasing function of λ

In fact, for $1 < p < \infty$, $F(\underset{\sim}{u}^\lambda)$ and $F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$ are increasing and $S(\underset{\sim}{u}^\lambda)$ is decreasing, provided that $S(\underset{\sim}{u}^\lambda) > 0$.

Furthermore, $(F(\underset{\sim}{u}^\lambda), S(\underset{\sim}{u}^\lambda))$ is *Pareto-optimal* (see Gerber [4], 91):

There does not exist $\underset{\sim}{u}$ such that

$F(\underset{\sim}{u}) \leqslant F(\underset{\sim}{u}^\lambda)$ and $S(\underset{\sim}{u}) \leqslant S(\underset{\sim}{u}^\lambda)$ with at least one inequality being strict.

Numerical examples are given in Section 5, in which graduated values obtained using different p are compared.

Lowrie [11] extended F and included exponential smoothness in S. In the discussion of Lowrie's paper, Chan, Chan and Mead [2] showed that the extension also has the Monotone Properties and is Pareto-optimal.

The Whittaker-Henderson graduation method [16] has a Bayesian statistical interpretation, which has been advanced by Hickman and Miller [7, 8].

Modifications of the graduation problem to the problems

$$\underset{\underset{\sim}{u}}{\text{Min}}\ F(\underset{\sim}{u}) \text{ under the constraint } S(\underset{\sim}{u}) \leqslant c$$

and

$$\underset{\underset{\sim}{u}}{\text{Min}}\ S(\underset{\sim}{u}) \text{ under the constraint } F(\underset{\sim}{u}) \leqslant c,$$

where $c > 0$ is a given constant, are given in Section 6.

Lemmas and theorems which require longer proofs are given in the appendices.

2. ℓ_p-NORM WITH $1 < p < \infty$

Let $1 < p < \infty$ and $\lambda \geqslant 0$ be given constants. Consider the following problem:

$$\underset{\underset{\sim}{u}}{\text{Min}}[F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u})] \tag{WH}$$

where

$$F(\underset{\sim}{u}) \equiv \sum_{x=1}^{n} w_i \mid u_x - u_x'' \mid^p \text{ and } S(\underset{\sim}{u}) \equiv \sum_{x=1}^{n-z} \mid \Delta^z u_x \mid^p.$$

We proceed to show that the optimal solution $\underset{\sim}{u}^\lambda = (u_1^\lambda, \ldots, u_n^\lambda)T$ to (WH) is the unique solution of the system of equations

$$F'(\underset{\sim}{u}) + \lambda S'(\underset{\sim}{u}) = \underset{\sim}{0}.^T \tag{2.1}$$

LEMMA 2.1: *$F(\underset{\sim}{u})$ and $S(\underset{\sim}{u})$ are differentiable at every point $\underset{\sim}{u} \in R^n$. They are twice differentiable except when $u_x = u_x''$ or $\Delta^z u_x = 0$ for some x for $1 < p < 2$, and*

(i)
$$F'(\underset{\sim}{u}) \equiv \left[\frac{\partial F}{\partial u_1}, \ldots, \frac{\partial F}{\partial u_n}\right] = p[|u_1 - u_1''|^{p-1}\,\mathrm{sgn}(u_1 - u_1''), \ldots, |u_n - u_n''|^{p-1}\,\mathrm{sgn}(u_n - u_n'')]W$$

$$S'(\underset{\sim}{u}) \equiv \left[\frac{\partial S}{\partial u_1}, \ldots, \frac{\partial S}{\partial u_n}\right] = p[|\Delta^z u_1|^{p-1}\mathrm{sgn}(\Delta^z u_1), \ldots, |\Delta^z u_{n-z}|^{p-1}\mathrm{sgn}(\Delta^z u_{n-z})]K,$$

where W is the $n \times n$ diagonal matrix with diagonal elements $w_1, \ldots, w_n$, and K is the $(n-z) \times n$ zth differencing matrix, i.e., $K\underset{\sim}{u}$ is the column vector $(\Delta^z u_1, \ldots, \Delta^z u_{n-z})^T$, *and*

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

(ii)
$$F''(\underset{\sim}{u}) \equiv \left[\frac{\partial^2 F}{\partial u_i \partial u_j}\right] = p\,(p-1) \begin{bmatrix} |u_1 - u_1''|^{p-2} & & 0 \\ & \ddots & \\ 0 & & |u_n - u_n''|^{p-2} \end{bmatrix} W$$

$$S''(\underset{\sim}{u}) \equiv \left[\frac{\partial^2 S}{\partial u_i \partial u_j}\right] = p\,(p-1)K^T \begin{bmatrix} |\Delta^z u_1|^{p-2} & & 0 \\ & \ddots & \\ 0 & & |\Delta^z u_{n-z}|^{p-2} \end{bmatrix} K$$

Proof. See Appendix I(a).

It is clear from (ii) that $S''(\underset{\sim}{u})$ and $F''(\underset{\sim}{u})$ are nonnegative definite if they exist. Therefore, $F''(\underset{\sim}{u}) + \lambda S''(\underset{\sim}{u})$ is nonnegative definite for every $\underset{\sim}{u}$, and hence $F + \lambda S$ is convex on R^n.

In fact, since $|x|^p$ is a strictly convex function, we have

LEMMA 2.2: *$F + \lambda S$ is strictly convex on R^n.*

Proof. See Appendix I(b).

Based on a documented property of a strictly convex function of several variables [3, §2.1], we have

THEOREM 2.1: *For a given* $\lambda \geq 0$, $\underset{\sim}{u}^{\lambda}$ *is an optimal solution to* (*WH*) *if and only if* $F'(\underset{\sim}{u}^{\lambda}) + \lambda S'(\underset{\sim}{u}^{\lambda}) = \underset{\sim}{0}$. *Furthermore, this optimal solution is unique.*

Since $F + \lambda S$ is strictly convex, the following Newton-Raphson algorithm [12, p.288] is used for solving the system of equations (2.1).

THEOREM 2.2: *Let* $\underset{\sim}{u}^{o}$ *be an initial value and* $\underset{\sim}{u}^{k}$ *denote the value after k iterations. If we set*

$$\underset{\sim}{u}^{k+1} = \underset{\sim}{u}^{k} - [F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})]^{-1}[F'(\underset{\sim}{u}^{k}) + \lambda S'(\underset{\sim}{u}^{k})]^{T},$$

and it converges to $\underset{\sim}{u}$, *then* $\underset{\sim}{u}$ *is the unique solution of equation* (2.1) *provided that for each k,* $[F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})]^{-1}$ *exists.*

When $p > 2$, $F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})$ is nonsingular if $F''(\underset{\sim}{u}^{k})$ is nonsingular, which can be achieved if $u_x^k \neq u_x''$ for all $x = 1, \ldots, n$. So, in case $u_x^k - u_x'' = 0$ for some x, we can always change u_x^k to $u_x^k + \epsilon$ with $\epsilon \neq \underset{\sim}{0}$.

For the case $p = 2$, Greville's [5] graduated values can be obtained immediately from Theorem 2.2. Since

$$[F'(\underset{\sim}{u}) + \lambda S'(\underset{\sim}{u})]^{T} = 2W(\underset{\sim}{u} - \underset{\sim}{u}'') + 2\lambda K^{T}K\underset{\sim}{u},$$

and

$$F''(\underset{\sim}{u}) + \lambda S''(\underset{\sim}{u}) = 2W + 2\lambda K^{T}K,$$

and the latter is positive definite and, hence, nonsingular, then for any $\underset{\sim}{u}^{o}$,

$$\begin{aligned}
\underset{\sim}{u}^{1} &= \underset{\sim}{u}^{o} - [F''(\underset{\sim}{u}^{o}) + \lambda S''(\underset{\sim}{u}^{o})]^{-1}[F'(\underset{\sim}{u}^{o}) + \lambda S'(\underset{\sim}{u}^{o})]^{T} \\
&= \underset{\sim}{u}^{o} - [2W + 2\lambda K^{T}K]^{-1}[2W(\underset{\sim}{u}^{o} - \underset{\sim}{u}'') - 2\lambda K^{T}K(\underset{\sim}{u}^{o})] \\
&= \underset{\sim}{u}^{o} - [W + \lambda K^{T}K]^{-1}[(W + \lambda K^{T}K)(\underset{\sim}{u}^{o}) - W\underset{\sim}{u}''] \\
&= \underset{\sim}{u}^{o} - \underset{\sim}{u}^{o} + (W + \lambda K^{T}K)^{-1}W\underset{\sim}{u}'' \\
&= (W + \lambda K^{T}K)^{-1}W\underset{\sim}{u}'',
\end{aligned}$$

which is independent of $\underset{\sim}{u}^{o}$.

For $1<p<2$, $u_x^k = u_x''$ or $\Delta^z u_x^k = 0$ for some x will lead to infinity in the entries of $F''(\underset{\sim}{u}^{k})$ or $S''(\underset{\sim}{u}^{k})$; that is, $F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})$ and $[F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})]^{-1}$ do not exist.

An APL program is written for carrying out the iterations in Theorem 2.2 and is given in Appendix I(b).

Usually $F''(\underset{\sim}{u}^{k}) + \lambda S''(\underset{\sim}{u}^{k})$ is a large matrix. Therefore, the square-root method or Choleski method (which can be found in Greville [5]) is recommended for finding $\underset{\sim}{u}^{k+1}$. We can find $\underset{\sim}{u}^{k+1}$ by using the above method to solve $[F''(\underset{\sim}{u}^k) + \lambda S''(\underset{\sim}{u}^k)]\ \underset{\sim}{u}^{k+1} = [F''(\underset{\sim}{u}^k) + \lambda S''(\underset{\sim}{u}^k)]\ \underset{\sim}{u}^k - [F'(\underset{\sim}{u}^k) + \lambda S'(\underset{\sim}{u}^k)]$ instead of inverting $F''(\underset{\sim}{u}^k) + \lambda S''(\underset{\sim}{u}^k)$.

3. ℓ_∞-NORM

In this section, we solve the ℓ_∞-norm problem raised by Greville [6] by formulating it as a linear programming problem.

The problem then is

$$\min_{\underset{\sim}{u} \geq \underset{\sim}{0}} [F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u})] \tag{WH}$$

where

$$F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} w_x |u_x - u_x''| \text{ and } S(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n-z} |\Delta^z u_x|.$$

We will consider the more general form of $F(\underset{\sim}{u})$, that is,

$$F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} w_x \mid u_x - u_x''|,$$

which has $F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} |u_x - u_x''|$ as a special case.

THEOREM 3.1: *Let λ and $w_1, \ldots, w_n$ be given constants. The Whittaker-Henderson graduation method with ℓ_∞ − norm* (WH)*and $\underset{\sim}{u} \geq \underset{\sim}{0}$ is equivalent to the linear programming problem, whose optimal solution always exists, of:*

$$\min_{\underset{\sim}{u}, f, s} (f + \lambda s) \tag{LP}$$

under the constraints

$$\left.\begin{aligned} w_x u_x - w_x u_x'' &\leq f, \\ -w_x u_x + w_x u_x'' &\leq f, \end{aligned}\right\} \quad x = 1, \ldots, n, \tag{3.1a}$$

$$\left.\begin{aligned} \sum_{i=1}^{n} u_i k_{xi} &\leq s, \\ -\sum_{i=1}^{n} u_i k_{xi} &\leq s, \end{aligned}\right\} \quad x = 1, \ldots, n-z, \tag{3.1b}$$

$$u_1, \ldots, u_n, f, s \geq 0,$$

where $[k_{xi}] = K$ *is the zth difference matrix. If* $(\underset{\sim}{u}^\lambda, f_\lambda, s_\lambda)$ *is an optimal solution to* (LP), *then* $\underset{\sim}{u}^\lambda$ *is an optimal solution to* (WH) *and*

$$f_\lambda = F(\underset{\sim}{u}^\lambda) \text{ and } s_\lambda = S(\underset{\sim}{u}^\lambda).$$

Proof. See Appendix II.

If the u_x are not constrained to be nonnegative in the (LP) formulation, they can be replaced by $u_x = u_x^* - u_x^{**}$, where $u_x^* \geq 0$ and $u_x^{**} \geq 0$.

Linear programming is the most widely used mathematical optimization model in operations research. Its optimal solution can be easily obtained by the simplex method. More about linear programming can be found in the Society of Actuaries Part 3 Examination reference, Hillier and Lieberman [9].

Since the optimal solution of (LP) and, hence, that of (WH) may not be unique, the first optimal solution obtained after the linear programming may not be a good fit to $\underset{\sim}{u}$. We can improve the goodness of fit of the solution by using quadratic programming. For example, suppose we have obtained $\underset{\sim}{u}^\lambda$ as an optimal solution and that $F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u}) = M$. We can then formulate the quadratic programming problem

$$\underset{\underset{\sim}{u}, f, s}{\text{Min}} \sum_{x=1}^{n} w_x(u_x - u_x'')^2 \qquad \text{(QP)}$$

under the constraints (3.1a), (3.1b) and the additional constraint

$$f + \lambda s \leq M.$$

It can be easily seen that any solution of (QP) is a solution of (WH).

However, for the case with $w_x = 1$ in $F(\underset{\sim}{u})$, the optimal solution obtained at the end of the linear programming is close to $\underset{\sim}{u}''$ (though it may not be unique), and no further quadratic programming is needed.

Another advantage of formulating (WH) as a linear programming problem is that one can perform *sensitivity analyses*. This is analyzing the effects on the optimal solution $(u_\lambda, f_\lambda, s_\lambda)$ of one or more of the following changes:

$$\lambda \rightarrow \lambda + \delta\lambda$$

$$u_x'' \rightarrow u_x'' + \delta u_x''$$

$$w_x \rightarrow w_x + \delta w_x.$$

The computational procedures for doing this analysis are given in [9, §5.3]. They are based on the final simplex tableau which usually can be obtained from the printout of a linear programming computer package.

4. MONOTONE PROPERTIES AND PARETO-OPTIMALITY

In this section general ℓ_p-norms, $1 \leq p \leq \infty$, are considered.

THEOREM 4:1: *The Monotone Properties hold and* $(F(\underset{\sim}{u}^\lambda), S(\underset{\sim}{u}^\lambda))$ *is Pareto-optimal.*

Proof. See Appendix III.

The Monotone Properties are intuitively clear. They can be used to check if some errors were made in the calculations of the $\underset{\sim}{u}^\lambda$ when several λ values are used. The case $p = 2$ was proven by Chan, Chan and Mead [1].

The Pareto-optimality says that $F(\underset{\sim}{u}^\lambda)$ and $S(\underset{\sim}{u}^\lambda)$ are the best values one can get. It is impossible to get a smaller value than $F(\underset{\sim}{u}^\lambda)$, or $S(\underset{\sim}{u}^\lambda)$, without getting a larger value than $S(\underset{\sim}{u}^\lambda)$, or $F(\underset{\sim}{u}^\lambda)$.

5. A NUMERICAL EXAMPLE

The $n = 19$ ungraduated values and weights given by Miller [13, p.35] are graduated using $p = 1,2,3,5$, and ∞ (with $F(\underset{\sim}{u}) = \max_{1 \leq x \leq n} |u_x - u''_x|$ and $F(\underset{\sim}{u}) = \max_{1 \leq x \leq n} w_x|u_x - u''_x|$) when $z = 3$ and $\lambda = 1, 2, 3, 6, 10$ (see Tables 1-6). The case $p = 10$ was also calculated; its graduated values are omitted because they are quite close to those when $p = 5$. Notice that the monotone properties are satisfied for all cases.

For the case $p = \infty$ with

$$F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} w_x|u_x - u''_x|,$$

some initial graduated values obtained at the end of the linear programming calculations were quite far from the ungraduated values. Much improvement was made after the quadratic programming calculations (see Table 6). For the case $p = \infty$ with

$$F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} |u_x - u''_x|,$$

the graduated values obtained at the end of the linear programming calculations are quite close to the ungraduated values. Therefore, we do not need quadratic programming to improve the fit (see Table 5). A graphical comparison of some graduation values is given in Figure 1 when $z = 3$ and $\lambda = 3$.

TABLE 1

GRADUATED VALUES WHEN $p = 1$ AND $z = 3$

x	Ungraduated Values u''_x	Weights w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated Values u^{λ}_x				
1....	34	3	34.00	34.00	34.00	15.90	22.32
2....	24	5	24.00	24.00	29.00	24.00	26.68
3....	31	8	31.00	31.00	31.00	31.00	31.00
4....	40	10	40.00	37.50	40.00	36.90	35.29
5....	30	15	30.00	43.50	46.00	41.70	39.56
6....	49	20	49.00	49.00	49.00	45.40	43.79
7....	48	23	48.00	48.00	48.00	48.00	48.00
8....	48	20	48.00	48.00	48.00	51.46	52.18
9....	67	15	67.00	51.67	51.67	55.78	56.73
10....	58	13	58.00	58.00	58.00	60.96	61.68
11....	67	11	67.00	67.00	67.00	67.00	67.00
12....	75	10	75.00	75.00	73.00	72.01	72.71
13....	76	9	76.00	76.00	76.00	76.00	78.80
14....	76	9	76.00	81.92	82.14	81.26	85.27
15....	102	7	102.00	92.75	91.43	87.79	92.13
16....	100	5	100.00	100.00	100.00	95.59	99.37
17....	101	5	101.00	103.67	107.86	104.66	107.00
18....	115	3	112.33	115.00	115.00	115.00	115.00
19....	134	1	134.00	134.00	121.43	126.61	123.39
Fit $F(\underset{\sim}{u}^{\lambda})$			8.01	588.83	691.07	833.23	873.01
Smoothness $S(\underset{\sim}{u}^{\lambda})$			415.34	76.03	35.41	6.15	0.63
$F(\underset{\sim}{u}^{\lambda}) + \lambda S(\underset{\sim}{u}^{\lambda})$			423.35	740.89	797.30	870.13	879.31

TABLE 2

GRADUATED VALUES WHEN $p = 2$ AND $z = 3$

x	Ungraduated Values u''_x	Weights w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated Values u^{λ}_x				
1 ..	34	3	31.65	31.17	30.94	30.58	30.30
2 ..	24	5	27.57	28.31	28.61	28.96	29.12
3 ..	31	8	30.98	30.76	30.68	30.64	30.69
4 ..	40	10	34.86	34.28	34.08	33.91	33.88
5 ..	30	15	35.95	36.93	37.33	37.76	37.93
6 ..	49	20	45.40	44.66	44.30	43.85	43.62
7 ..	48	23	48.16	48.21	48.25	48.30	48.33
8 ..	48	20	51.38	52.10	52.44	52.87	53.09
9 ..	67	15	61.04	59.98	59.53	58.99	58.73
10 ..	58	13	62.19	62.68	62.83	62.90	62.88
11 ..	67	11	66.86	67.00	67.05	67.10	67.11
12 ..	75	10	72.65	72.06	71.86	71.72	71.73
13 ..	76	9	75.63	75.98	76.21	76.58	76.81
14 ..	76	9	81.75	82.60	82.94	83.30	83.44
15 ..	102	7	94.76	93.53	92.93	92.10	91.66
16...	100	5	100.69	100.11	99.80	99.37	99.13
17 ..	101	5	104.18	105.08	105.55	106.20	106.53
18 ..	115	3	114.00	114.55	114.89	115.40	115.68
19 ..	134	1	132.07	130.36	129.38	127.98	127.25
Fit $F(\underset{\sim}{u}^{\lambda})$			2,905.68	3,980.60	4,502.81	5,164.97	5,488.96
Smoothness $S(\underset{\sim}{u}^{\lambda})$			1,233.80	451.84	236.04	73.14	30.15
$F(\underset{\sim}{u}^{\lambda}) + \lambda S(\underset{\sim}{u}^{\lambda})$			4,139.48	4,884.29	5,210.92	5,603.83	5,790.45

TABLE 3

GRADUATED VALUES WHEN $p = 3$ AND $z = 3$

x	Ungraduated Values u''_x	Weights w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated Values u^λ_x				
1 ..	34	3	30.91	30.71	30.60	30.42	30.29
2 ..	24	5	28.00	28.24	28.36	28.53	28.64
3 ..	31	8	30.97	30.73	30.63	30.51	30.45
4 ..	40	10	34.46	34.14	34.00	33.81	33.71
5 ..	30	15	36.14	36.53	36.72	36.98	37.13
6 ..	49	20	44.31	43.98	43.81	43.57	43.43
7 ..	48	23	48.43	48.43	48.44	48.46	48.47
8 ..	48	20	52.64	52.98	53.16	53.41	53.57
9 ..	67	15	60.95	60.56	60.37	60.09	59.92
10 ..	58	13	62.82	63.07	63.18	63.31	63.38
11 ..	67	11	66.39	66.46	66.51	66.60	66.72
12 ..	75	10	71.39	71.09	70.96	70.85	70.83
13 ..	76	9	74.72	74.97	75.13	75.44	75.68
14 ..	76	9	82.24	82.74	82.99	83.33	83.52
15 ..	102	7	94.92	94.34	94.03	93.56	93.25
16 ..	100	5	101.62	101.22	101.00	100.69	100.51
17 ..	101	5	105.46	105.93	106.19	106.60	106.88
18 ..	115	3	113.64	114.22	114.57	115.15	115.51
19 ..	134	1	130.07	129.36	129.03	128.55	128.22
Fit $F(\underline{u}^\lambda)$			20,117.30	24,600.39	27,080.02	30,854.36	33,295.22
Smoothness $S(\underline{u}^\lambda)$			5,832.85	2,593.14	1,572.36	656.18	335.08
$F(\underline{u}^\lambda) + \lambda S(\underline{u}^\lambda)$			25,950.15	29,786.68	31,797.11	34,791.46	36,646.05

TABLE 4

GRADUATED VALUES WHEN $p = 5$ AND $z = 3$

x	Ungraduated Values u''_x	Weights w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated Values u^λ_x				
1 ..	34	3	30.12	30.03	29.99	29.92	29.89
2 ..	24	5	28.49	28.60	28.65	28.72	28.76
3 ..	31	8	31.60	31.45	31.36	31.22	31.12
4 ..	40	10	34.33	34.19	34.11	33.99	33.92
5 ..	30	15	36.10	36.27	36.36	36.50	36.59
6 ..	49	20	43.63	43.47	43.39	43.26	43.17
7 ..	48	23	48.56	48.57	48.57	48.54	48.51
8 ..	48	20	53.33	53.48	53.56	53.70	53.79
9 ..	67	15	60.99	60.82	60.73	60.60	60.52
10 ..	58	13	63.27	63.42	63.49	63.58	63.61
11 ..	67	11	66.63	66.90	66.97	66.94	66.84
12 ..	75	10	70.19	70.02	69.94	69.83	69.78
13 ..	76	9	73.31	73.42	73.51	73.74	73.95
14 ..	76	9	82.47	82.70	82.83	83.03	83.17
15 ..	102	7	95.14	94.88	94.73	94.49	94.31
16 ..	100	5	102.57	102.30	102.15	101.93	101.79
17 ..	101	5	106.41	106.65	106.78	107.01	107.17
18 ..	115	3	113.45	113.90	114.15	114.56	114.83
19 ..	134	1	128.85	128.61	128.47	128.25	128.09
Fit $F(\underline{u}^\lambda)$			805,039	938,030	1,016,166	1,148,252	1,243,603
Smoothness $S(\underline{u}^\lambda)$			189,265	94,092	62,015	38,312	17,718
$F(\underline{u}^\lambda) + \lambda S(\underline{u}^\lambda)$			994,904	1,126,215	1,202,211	1,378,123	1,420,780

TABLE 5

GRADUATED VALUES WHEN $p = \infty$, $z = 3$

x	Ungraduated Values u''_x	Weights* w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated values u^λ_x				
1....	34	1	24.94	24.81	24.56	24.39	24.39
2....	24	1	27.07	27.41	27.68	27.90	27.91
3....	31	1	30.32	30.79	31.26	31.64	31.65
4....	40	1	34.41	34.78	35.21	35.55	35.56
5....	30	1	39.06	39.19	39.44	39.61	39.61
6....	49	1	43.99	43.86	43.86	43.79	43.79
7....	48	1	48.93	48.60	48.38	48.12	48.12
8....	48	1	53.58	53.26	52.92	52.64	52.64
9....	67	1	57.94	57.81	57.56	57.39	57.39
10....	58	1	62.28	62.44	62.40	62.39	62.39
11....	67	1	66.89	67.32	67.51	67.68	67.68
12....	75	1	72.03	72.62	73.00	73.29	73.29
13....	76	1	77.99	78.52	78.94	79.26	79.26
14....	76	1	85.06	85.19	85.44	85.61	85.61
15....	102	1	92.94	92.81	92.56	92.39	92.39
16....	100	1	101.37	101.20	100.24	99.62	99.62
17....	101	1	110.06	110.19	108.37	107.34	107.34
18....	115	1	118.73	119.60	116.87	115.58	115.59
19....	134	1	127.66	129.62	125.66	124.38	124.39
Fit $F(\underset{\sim}{u}^\lambda)$			9.06	9.19	9.44	9.62	9.61
Smoothness $S(\underset{\sim}{u}^\lambda)$			0.30	0.19	0.11	0.06	0.05
$F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$			9.36	9.57	9.77	9.98	10.11

TABLE 6

GRADUATED VALUES WHEN $p = \infty$ AND $z = 3$

x	Ungraduated Values u''_x	Weights* w_x	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 6$	$\lambda = 10$
			Graduated Values u^λ_x				
1....	34	3	34.00	34.00	27.11	27.07	16.40
2....	24	5	24.00	24.00	24.12	24.10	21.75
3....	31	8	31.00	31.00	25.85	25.84	27.10
4....	40	10	39.14	39.14	30.72	30.72	32.45
5....	30	15	30.57	30.58	37.17	37.17	37.80
6....	49	20	48.57	48.57	43.62	43.62	43.15
7....	48	23	48.37	48.37	48.50	48.50	48.50
8....	48	20	48.35	48.35	53.38	53.38	53.85
9....	67	15	66.54	66.54	59.83	59.83	59.20
10....	58	13	58.18	58.18	66.28	66.28	64.55
11....	67	11	67.00	67.00	71.15	71.16	69.90
12....	75	10	75.00	75.00	74.60	74.61	75.25
13....	76	9	75.04	75.04	78.21	78.22	80.60
14....	76	9	76.96	76.96	83.54	83.54	85.95
15....	102	7	100.77	100.77	90.85	90.85	91.30
16....	100	5	101.72	101.73	98.57	98.57	96.65
17....	101	5	101.00	101.00	106.35	106.36	102.00
18....	115	3	115.00	115.00	115.77	115.77	107.35
19....	134	1	134.00	134.00	128.39	128.38	112.70
Fit $F(\underset{\sim}{u}^\lambda)$			8.64	8.70	107.64	107.64	117.00
Smoothness $S(\underset{\sim}{u}^\lambda)$			44.77	44.76	1.58	1.59	0.00
$F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$			53.41	98.22	112.38	117.18	117.00

* $F(\underset{\sim}{u}) \equiv \max_{1 \le x \le 19} w_x \mid u_x - u''_x \mid$; the $F(\underset{\sim}{u}^\lambda)$ of Table 5 is the special case when all $w_x = 1$.

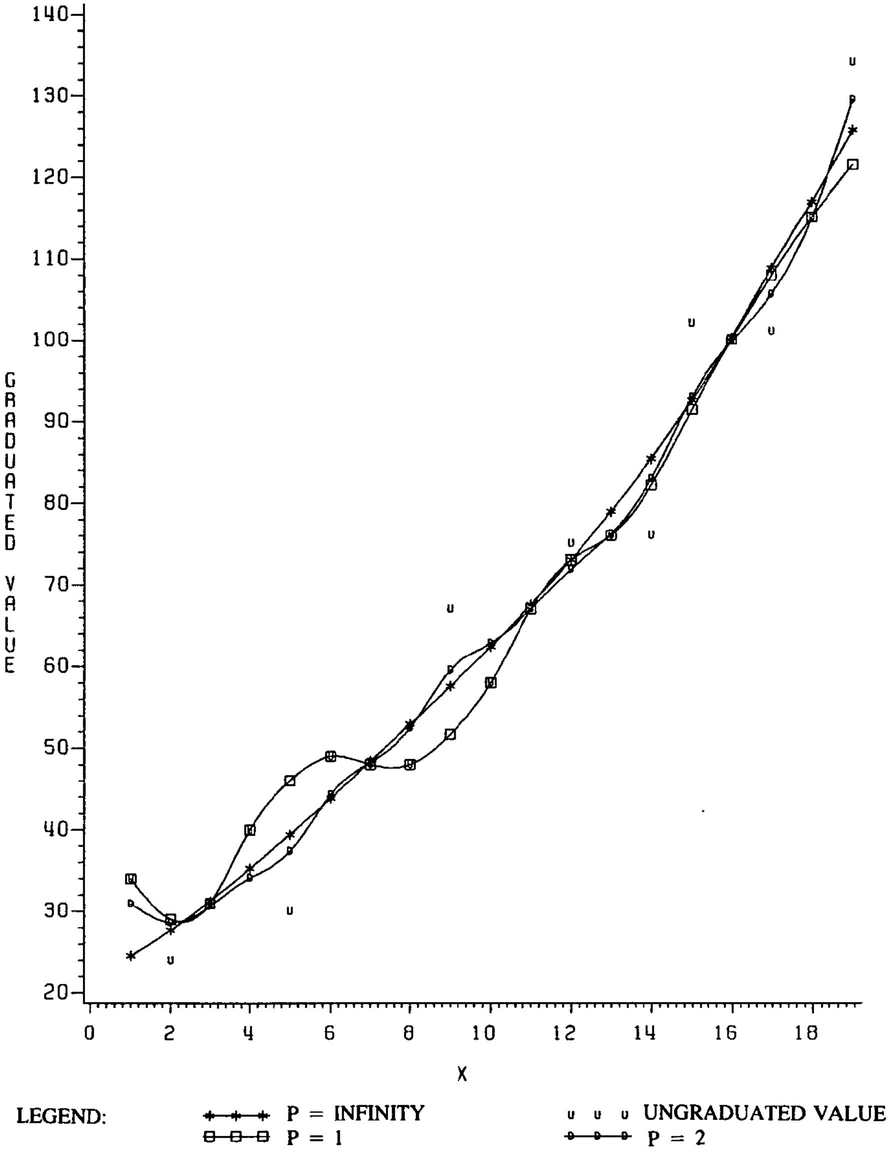

FIG. 1—Comparison of Graduated Values, Z = 3 and λ = 3

6. MODIFICATIONS OF THE WHITTAKER-HENDERSON GRADUATION METHOD

The traditional approach of minimizing $F + \lambda S$ is now modified to the minimization of F under the constraint that S does not exceed a predetermined value c.

(i) ℓ_∞ – norm case: The problem

$$\underset{\underset{\sim}{u} \geq 0}{\text{Min}}[\max_{1 \leq x \leq n} w_x \mid u_x - u''_x \mid]$$

under the constraints

$$\max_{1 \leq x \leq n-z} \mid \Delta^z u_x \mid \leq c$$

is equivalent to the linear programming problem

$$\min_{\underset{\sim}{u} \geq \underset{\sim}{0}} f$$

under the constraints (3.1a) and (3.1b) with s replaced by c. The value c should be chosen such that $c \leq S(\underset{\sim}{u}'')$ since $F(\underset{\sim}{u}'') = 0$.

(ii) ℓ_1 – norm case: Schuette [14] formulated the ℓ_1 – norm case as the linear program problem

$$\text{Min} \sum_{x=1}^{n} w_x(P_x + N_x) + \lambda \sum_{x=1}^{n-z} (R_x + T_x)$$

under the constraints

$$K(P - N) + I_{n-z}(R - T) = K\underset{\sim}{u}'',$$

where $\underset{\sim}{u}'' - \underset{\sim}{u} = P - N$, $\Delta^z \underset{\sim}{u} = R - T$ with $P_x, N_x, R_x, T_x \geq 0$, and I_{n-z} is the identity matrix of order $n - z$.

The minimization of F under $S \leq c$ can be formulated as

$$\text{Min} \sum_{x=1}^{n} w_x(P_x + N_x)$$

under the constraints

$$K(P-N) + I_{n-z}(R-T) = K\underset{\sim}{u}'' \text{ and}$$

$$\sum_{x=1}^{n-z} (R_x + T_x) \leq c.$$

(iii) ℓ_p-norm case, $1 < p < \infty$: The problem is

$$\underset{\underset{\sim}{u}}{\text{Min}} \sum_{x=1}^{n} w_x |u_x - u_x''|^p$$

under the constraints

$$\sum_{x=1}^{n-z} |\Delta^z u_x|^p \leq c,$$

where

$$S(\underset{\sim}{u}'') \geq c.$$

Using a proof similar to that of Theorem 2 of Chan, Chan and Mead [1], one can show that the optimal solution is the unique solution to the system of $n+1$ equations

$$F'(\underset{\sim}{u}) + \beta S'(\underset{\sim}{u}) = \underset{\sim}{0}^T,$$

$$S(\underset{\sim}{u}) = c,$$

where β is the Lagrange multiplier.

The problem of minimization of S under the constraint that $F \leq c$ for the ℓ_p -norm cases, $1 \leq p \leq \infty$, can be similarly formulated and solved.

7. CONCLUSION

This paper considers the Whittaker-Henderson graduation method with general ℓ_p -norm, $1 \leq p \leq \infty$. It has been shown that for $1 < p < \infty$, the set of graduated values $\underset{\sim}{u}^\lambda = (u_1^\lambda, \ldots, u_n^\lambda)$ is unique and is the solution of the system of equations $F'(\underset{\sim}{u}) + \lambda S'(\underset{\sim}{u}) = \underset{\sim}{0}^T$.

When $p = \infty$, the method can be formulated as a linear programming problem. With $F(\underset{\sim}{u}) \equiv \max_{1 \leq x \leq n} w_x |u_x - u_x''|$, the optimal solution obtained at the end of the linear programming may not be unique and could be quite far from the ungraduated values $\underset{\sim}{u}''$. The fit can be improved through quadratic programming. However, with $F(\underset{\sim}{u}) = \max_{1 \le x \le n} |u_x - u''_x|$, no quadratic programming is required.

For $1 \le p \le \infty$, it is shown that $F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda)$, $F(\underset{\sim}{u}^\lambda)$, and $S(\underset{\sim}{u}^\lambda)$ are, respectively, nondecreasing, nondecreasing and nonincreasing functions of λ, and that there does not exist a $\underset{\sim}{u}$ such that $F(\underset{\sim}{u}) \le F(\underset{\sim}{u}^\lambda)$ and $S(\underset{\sim}{u}) \le S(\underset{\sim}{u}^\lambda)$ with at least one inequality being strict.

When $1 \le p \le \infty$, it is shown that the alternative of minimizing F (or S) subject to $S \le c$ (or $F \le c$) has some of its properties and solution algorithms analogous to the traditional method of minimizing $F + \lambda S$.

APPENDIX I(a)

Here, $1 < p < \infty$. Let $G: R \to R$ with $G(y) = |y|^p$. Then

$$G'(y) = p\, sgn(y)|y|^{p-1}$$

for every y and p, and

$$G''(y) = p(p-1)|y|^{p-2}$$

except when $y = 0$ and $p < 2$ [3, p.26]. For a function $M: R^r \to R^s$, define

$$M'(\underset{\sim}{y}) \equiv \begin{bmatrix} \frac{\partial M_1}{\partial y_1} & \cdots & \frac{\partial M_1}{\partial y_r} \\ \cdots & \cdots & \cdots \\ \frac{\partial M_s}{\partial y_1} & \cdots & \frac{\partial M_s}{\partial y_r} \end{bmatrix}.$$

If

$$M = \begin{bmatrix} m_{11} & \cdots & m_{1r} \\ \cdots & \cdots & \cdots \\ m_{s1} & \cdots & m_{sr} \end{bmatrix},$$

define $M(\underset{\sim}{y}) \equiv M\, \underset{\sim}{y}$ with $y \in R^r$. Then $M'(y) = M$.

If $A: R^r \to R^s$ and $B: R^s \to R^t$, and $B(A)$ is the composite function: $R^r \to R^t$, then the Chain Rule is

$$C'(\underset{\sim}{y}) = B'(A(\underset{\sim}{y}))\cdot A'(\underset{\sim}{y}),$$

where

$$C(\underset{\sim}{y}) = B(A(\underset{\sim}{y})).$$

Proof of Lemma 2.1:

(*i*) $F(\underset{\sim}{u})$ can be expressed as $B(A(\underset{\sim}{u}))$, where

$$A(\underset{\sim}{u}) = \begin{bmatrix} |u_1 - u_1''|^p \\ \cdot \\ \cdot \\ \cdot \\ |u_n - u_n''|^p \end{bmatrix}, \ B(\underset{\sim}{y}) = \sum_{x=1}^{n} w_x y_x.$$

Then

$$A'(\underset{\sim}{u}) = \begin{bmatrix} p|u_1 - u_1''|^{p-1} sgn(u_1 - u_1'') & & 0 \\ & \ddots & \\ 0 & & p|u_n - u_n''|^{p-1} sgn(u_n - u_n'') \end{bmatrix},$$

$$B'(\underset{\sim}{y}) = (w_1, \ldots, w_n).$$

So, by the Chain Rule,

$$\begin{aligned} F'(\underset{\sim}{u}) &= B'(A(\underset{\sim}{u}))A'(\underset{\sim}{u}) \\ &= [p|u_1 - u_1''|^{p-1} sgn(u_1 - u_1''), \ldots, p|u_n - u_n''|^{p-1} sgn(u_n - u_n'')]W. \end{aligned}$$

$S(u)$ can be expressed as $B(K(u))$, where K is the $(n-z) \times n$ zth differencing matrix and $B(y) = \sum_{x=1}^{n-z} |y_x|^p$ with $\underset{\sim}{y} \in R^{n-z}$. Then $K'(\underset{\sim}{u}) = K$ and $B'(\underset{\sim}{y}) = [p\,(sgn\,(y_1))|y_1|^{p-1}, \ldots, p\,(sgn\,(y_{n-z}))|y_{n-z}|^{p-1}]$. So, by the Chain Rule, $S'(\underset{\sim}{u}) = [p|\Delta^z u_1|^{p-1} sgn\,(\Delta^z u_1), \ldots, p|\Delta^z u_{n-z}|^{p-1} sgn\,(\Delta^z u_{n-z})]K$.

(*ii*) The matrix $(F'(\underset{\sim}{u}))^T$ can be expressed as $p(W(A))(\underset{\sim}{u})$, where

$$A(\underset{\sim}{u}) = \begin{bmatrix} |u_1 - u_1''|^{p-1} sgn(u_1 - u_1'') \\ \cdot \\ \cdot \\ \cdot \\ |u_n - u_n''|^{p-1} sgn(u_n - u_n'') \end{bmatrix}.$$

Then, by the Chain Rule,

$$F''(\underset{\sim}{u}) = p(p-1)WA'(\underset{\sim}{u}) = p(p-1)A'(\underset{\sim}{u})W,$$

where

$$A'(\underset{\sim}{u}) = \begin{bmatrix} |u_1 - u_1''|^{p-2} & & 0 \\ & \ddots & \\ 0 & & |u_n - u_n''|^{p-2} \end{bmatrix}.$$

The expression for $S''(\underset{\sim}{u})$ can be similarly derived.

Proof of Lemma 2.2:

For $1 < p < \infty$, $G'(x) = p\, sgn|x|^{p-1}$ is increasing. So $G(x) = |x|^p$ is strictly convex; for example, if $x^* \neq x$, then for $0 < \theta < 1$

$$|\theta x^* + (1-\theta)x|^p < \theta|x^*|^p + (1-\theta)|x|^p.$$

Therefore, if $\underset{\sim}{u}^\circ \neq \underset{\sim}{u}$, then for $0 < \theta < 1$,

$$\begin{aligned} F(\theta \underset{\sim}{u}^* + (1-\theta)\underset{\sim}{u}) &= \sum_{i=1}^{n} w_x|\theta u_x^* + (1-\theta)u_x - u_x''|^p \\ &= \sum_{i=1}^{n} w_x|\theta(u_x^* - u_x'') + (1-\theta)(u_x - u_x'')|^p \\ &< \theta \sum_{i=1}^{n} w_x|u_x^* - u_x''|^p + (1-\theta) \sum_{i=1}^{n} w_x|u_x - u_x''|^p \\ &= F(\underset{\sim}{u}^*) + (1-\theta)F(\underset{\sim}{u}), \end{aligned}$$

that is, F is strictly convex. Furthermore, S is convex since S'' is nonnegative definite. Consequently, $F + \lambda S$ is strictly convex.

APPENDIX I(b)

The following APL program for carrying out the iterations in Theorem 2.2 is illustrated by the numerical example (with $p = 5$, $z = 3$, $\lambda = 6$) in Section 5.

```
      ∇GRAD[⎕]∇
      ∇ GRAD IV

[1]   CR←1
[2]   →0×⍳0=CR
[3]   FF←P×(W+ . × (×IV−UV)×(|IV
      −UV)*P−1)+L×(⍉K)+.×(×K+.×IV)×(|K+.×IV)*P−1
[4]   A←19 19 ρ, (⍉(|IV-UV)*P−2), 19, 19 ρ0
[5]   B←Jρ,(⍉(|K+.×IV)*P−2),Jρ0
[6]   FFF←P × (P−1) × (W+ . ×A)+(L×(⍉K)+ . ×B+ . × K)
[7]   D←(⌹FFF)+.×FF
[8]   GV←IV←IV+0.1×UV=IV←IV−D
[9]   F←+/W+.×(|IV−UV)*P
[10]  S←+/(|K+.×IV)*P
[11]  M←F+L×S
[12]  CR←(⌈/|D)≥0.00001
[13]  →2
      ∇
      P←5
      L←6
      J←16 16
      UV←34 24 31 40 30 49 48 48 67 58 67 75 76 76 102 101 100 115 134
```

```
IV←36 22 33 38 32 47 50 46 69 56 69 73 78 74 104 98 103 113 136
K←16 19ρ¯1 3 ¯3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
W←19 19ρ, (19 1ρ3 5 8 10 15 20 23 20 15 13 11 10 9 9 7 5 5 3 1), 19 19ρ0

GRAD IV


GV
29.92     28.72     31.22     33.99     36.50
43.26     48.54     53.70     60.60     63.58
66.94     69.83     73.74     83.03
94.49     101.93    107.01    114.56
128.25
F
1148648
S
20156
F+L×S
1329589
```

Explanation of Symbols

P	= NORM	W	= WEIGHT MATRIX
L	= LAMBDA	GV	= GRADUATED VALUES
UV	= UNGRADUATED VALUES	F	= FIT
IV	= INITIAL ITERATION VALUES	S	= SMOOTHNESS
K	= ZTH DIFFERENCE MATRIX		

APPENDIX II

Proof of Theorem 3.1:

The (LP) problem has at least one feasible solution if f and s are large enough. Furthermore, an optimal solution to (LP) always exists because the objective function, which is to be minimized, is bounded below [9, p.97].

If $(\underset{\sim}{u}^\lambda, f_\lambda, s_\lambda)$ is an optimal solution to (LP), then by (3.1a) and (3.1b)

$$F(\underset{\sim}{u}^\lambda) \le f_\lambda \text{ and } S(\underset{\sim}{u}^\lambda) \le s_\lambda.$$

Suppose that one of the above inequalities, say the first inequality, is a strict inequality. Then $F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda) < f_\lambda + \lambda s_\lambda$. Since

$$F(\underset{\sim}{u}^\lambda) = \max_{1\le x\le n} w_x|u_x - u''_x| \ge w_x|u^\lambda_x - u'_x| \text{ for } x = 1,\ldots,n,$$

(3.1a) is satisfied with $f = F(\underset{\sim}{u}^\lambda)$. This is contradictory to $(\underset{\sim}{u}^\lambda, f_\lambda, s_\lambda)$ being optimal. So

$$f_\lambda = F(\underset{\sim}{u}^\lambda) \text{ and } s_\lambda = S(\underset{\sim}{u}^\lambda).$$

If $\underset{\sim}{u}_o$ is an optimal solution to (WH), that is,

$$f_o + \lambda s_o = \underset{\underset{\sim}{u} \geq \underset{\sim}{0}}{\text{Min}}[F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u})],$$

where $f_o \equiv F(\underset{\sim}{u}^o)$ and $s_o \equiv S(\underset{\sim}{u}^o)$, then

$$w_x|u_x^o - u_x''| \leq f_o \text{ for } x = 1,\ldots,n,$$

$$|\Delta^z u_x^o| \leq s_o \text{ for } x = 1,\ldots,n-z,$$

and, hence, $\underset{\sim}{u}^o$ satisfies the constraints (3.1a) and (3.1b) of (LP). Since $(\underset{\sim}{u}^\lambda, f_\lambda, s_\lambda)$ is an optimal solution to (LP),

$$f_\lambda + \lambda s_\lambda \leq f_o + \lambda s_o.$$

But the minimization in (WH) is over all $\underset{\sim}{u}$,

$$f_\lambda + \lambda s_\lambda \geq f_o + \lambda s_o.$$

Therefore

$$f_\lambda + \lambda s_\lambda = f_o + \lambda s_o,$$

and $\underset{\sim}{u}^\lambda$ is an optimal solution to (WH).

APPENDIX III

Proof of Theorem 4.1:

If $\lambda > \lambda^* \geq 0$, then

$$F(\underset{\sim}{u}^{\lambda^*}) + \lambda^* S(\underset{\sim}{u}^{\lambda^*}) \leq F(\underset{\sim}{u}^\lambda) + \lambda^* S(\underset{\sim}{u}^\lambda) \leq F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda). \tag{III.1}$$

The second inequality is strict if $S(\underset{\sim}{u}^\lambda) > 0$ and, hence, the first Monotone Property holds.

By adding the first inequality in (III.1) to the similar inequality

$$F(\underset{\sim}{u}^\lambda) + \lambda S(\underset{\sim}{u}^\lambda) \leq F(\underset{\sim}{u}^{\lambda^*}) + \lambda S(\underset{\sim}{u}^{\lambda^*}),$$

we obtain

$$0 \leq (\lambda - \lambda^*)\,[S(\underset{\sim}{u}^{\lambda^*}) - S(\underset{\sim}{u}^\lambda)].$$

That is, $S(\underset{\sim}{u}^\lambda) \leq S(\underset{\sim}{u}^{\lambda^*})$. This and (III.1) imply that $F(\underset{\sim}{u}^{\lambda^*}) \leq F(\underset{\sim}{u}^\lambda)$.

Now we proceed to show that for $1 < p < \infty$, $S(\underset{\sim}{u}^\lambda) < S(\underset{\sim}{u}^{\lambda^*})$ if $S(\underset{\sim}{u}^\lambda) > 0$. This holds if the first inequality in (III.1) is strict. If this is not true, then

$$F(\underset{\sim}{u}^{\lambda^*}) + \lambda^* S(\underset{\sim}{u}^{\lambda^*}) = F(\underset{\sim}{u}^\lambda) + \lambda^* S(\underset{\sim}{u}^\lambda),$$

and $\underset{\sim}{u}^{\lambda^*} \neq \underset{\sim}{u}^\lambda$, implying that the optimal solution to (WH) is not unique. This contradicts

Theorem 2.1. Consider the fact that $\underset{\sim}{u}^{\lambda^*} \neq \underset{\sim}{u}^{\lambda}$ comes from $F'(\underset{\sim}{u}^{\lambda}) + \lambda S'(\underset{\sim}{u}^{\lambda}) = \underset{\sim}{0} = F'(\underset{\sim}{u}^{\lambda^*}) + \lambda^* S'(\underset{\sim}{u}^{\lambda^*})$ (see Theorem 2.1) and the assumption that $S(\underset{\sim}{u}^{\lambda})>0$. Suppose $\underset{\sim}{u}^{\lambda^*} = \underset{\sim}{u}^{\lambda}$; then we have $\lambda^* S'(\underset{\sim}{u}^{\lambda^*}) = \lambda S'(\underset{\sim}{u}^{\lambda})$, which implies that $S'(\underset{\sim}{u}^{\lambda}) = S'(\underset{\sim}{u}^{\lambda^*}) = \underset{\sim}{0}$. Since

$$S'(\underset{\sim}{u}^{\lambda}) = p[|\Delta^z u_1^{\lambda}|^{p-1}\,\mathrm{sgn}(\Delta^z u_1^{\lambda}),\ldots,|\Delta^z u_{n-z}^{\lambda}|^{p-1}\,\mathrm{sgn}(\Delta^z u_{n-z}^{\lambda})]K,$$

$S'(\underset{\sim}{u}^{\lambda}) = \underset{\sim}{0}$ implies that $|\Delta^z u_x^{\lambda}| = 0$ for all $1 \leq x \leq n-z$; that is, $S(\underset{\sim}{u}^{\lambda}) = \underset{\sim}{0}$, which contradicts the assumption.

The first inequality in (III.1) and $S(\underset{\sim}{u}^{\lambda})<S(\underset{\sim}{u}^{*})$ implies that $F(\underset{\sim}{u}^{\lambda})>F(\underset{\sim}{u}^{\lambda^*})$.

To see that $(F(\underset{\sim}{u}_{\lambda}), S(\underset{\sim}{u}_{\lambda}))$ is Pareto-optimal, suppose that $\underset{\sim}{u}$ is such that $F(\underset{\sim}{u}) < F(\underset{\sim}{u}^{\lambda})$, and that $S(\underset{\sim}{u}) \leq S(\underset{\sim}{u}^{\lambda})$; then

$$F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u}) < F(\underset{\sim}{u}^{\lambda}) + \lambda S(\underset{\sim}{u}^{\lambda}),$$

which contradicts the assertion that $\underset{\sim}{u}^{\lambda}$ minimizes $F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u})$.

REFERENCES

1. CHAN, F.Y., CHAN, L.K., AND MEAD, E.R. "Properties and Modifications of Whittaker-Henderson Graduation," *Scandinavian Actuarial Journal*, 1982, 57–61.
2. CHAN, F.Y., CHAN, L.K., AND MEAD, E.R. Discussion on "An Extension of the Whittaker-Henderson Method of Graduation," *TSA*, XXXIV (1982), 368–369.
3. FLEMMING, W.H. *Functions of Several Variables*, Reading, MA: Addison-Wesley, 1965.
4. GERBER, H.U. *An Introduction to Mathematical Risk Theory*, Homewood, IL: Irwin, 1980.
5. GREVILLE, T.N.E. "Part 5 Study Notes—Graduation," Chicago, IL: Society of Actuaries, 1974.
6. GREVILLE, T.N.E. Discussion on "A Linear Programming Approach to Graduation," *TSA*, XXX (1978), 442.
7. HICKMAN, J.C., AND MILLER, R.B. "Notes on Bayesian Graduation," *TSA*, XXIX (1977), 7–21.
8. HICKMAN, J.C., AND MILLER, R.B. "Bayesian Bivariate Graduation and Forecasting," *Scandinavian Actuarial Journal*, 1981, 129–150.
9. HILLIER, F., AND LIEBERMAN, G. *Introduction to Operations Research* (3rd Edition), San Francisco, CA: Holden-Day, 1980.
10. HOFFMAN, K. *Analysis in Euclidean Space*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975.
11. LOWRIE, W.B. "An Extension of the Whittaker-Henderson Method of Graduation," *TSA* XXXIV (1982), 329–72.
12. MORDECAI, A. *Nonlinear Programming*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1976.

13. MILLER, M.D. *Elements of Graduation,* New York, NY: Actuarial Society of American and American Institute of Actuaries, 1946.
14. SCHUETTE, D.R. "A Linear Programming Approach to Graduation," *TSA*, XXX (1978), 407–445.
15. SHIU, E.S.W. "Matrix Whittaker-Henderson Graduation Formula," *ARCH*, Chicago, IL: Society of Actuaries, (1977), Issue No. 1.
16. WHITTAKER, E.T., AND ROBINSON, G., *The Calculus of Observations*, London: Blackie and Son, Ltd., 1924.

DISCUSSION OF PRECEDING PAPER

ELIAS S. W. SHIU:

This paper is an interesting extension of Schuette [7]. The results here can be generalized to the case where F and S are formulated by different ℓ_p-norms. The p for F should be small to diminish the effects of the outliers. The p for S should be large so that the graduated sequence is uniformly smooth. Thus, the problem is to minimize

$$\sum_x w_x|u_x - u_x''| + \lambda \max_y|\Delta^z u_y|.$$

For elaboration on the above, see [7], pages 434-45.

Indeed, the concept of Whittaker-Henderson graduation can be further generalized as: Find $\boldsymbol{u}$ which minimizes

$$h(\boldsymbol{u}) = f(\boldsymbol{u} - \boldsymbol{u}'') + g(K\boldsymbol{u}),$$

where f and g are convex functions. In the case where

$$f(\boldsymbol{x}) = \boldsymbol{x}^T W \boldsymbol{x}$$

and

$$g(\boldsymbol{x}) = \lambda\, \boldsymbol{x}^T \boldsymbol{x},$$

we have the classical Whittaker-Henderson type-B graduation. Let us assume that the convex functions f and g are twice-differentiable. Then, it may be possible to solve the equation

$$h'(\boldsymbol{u}) = \boldsymbol{0}^T$$

by Newton-Raphson iterations

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - [h''(\boldsymbol{u}^k)]^{-1}[h'(\boldsymbol{u}^k)]^T,$$

where

$$h'(\boldsymbol{u}) = f'(\boldsymbol{u}-\boldsymbol{u}'') + g'(K\boldsymbol{u})K$$

and

$$h''(\boldsymbol{u}) = f''(\boldsymbol{u}-\boldsymbol{u}'') + K^T g''(K\boldsymbol{u})K.$$

Note that the matrix K need not be a differencing matrix; see Greville [3], page 389.

In formulating the minimization problem, one should also specify the constraints

$$0 \leq u_x \leq 1,$$

if the u_x's are probabilities. Furthermore, linear constraints such as

$$c(x)u_x + d(x) \leq u_{x+1}$$

may be imposed on the graduated values as desired. The problem becomes one of minimizing

$$h(\boldsymbol{u}),\ \boldsymbol{u} \epsilon C, \tag{1}$$

where C is a closed and convex subset of R^n. For a differentiable function h, a vector $\boldsymbol{r} \in C$ satisfies

$$h(\boldsymbol{r}) = \operatorname*{Min}_{\boldsymbol{u} \epsilon C} h(\boldsymbol{u}) \tag{2}$$

only if

$$h'(\boldsymbol{r})\ (\boldsymbol{u} - \boldsymbol{r}) \geq 0 \quad \text{for each } \boldsymbol{u} \in C. \tag{3}$$

In general, (3) does not imply (2) unless h is convex (or pseudo-convex).

Since minimizing a nonlinear function of many variables under constraints is computationally complex, how does one minimize (1)? A solution has been forwarded eloquently by W. Conley [2]:

> Computer technology has advanced to the point that it is now possible to take an entirely different philosophical approach to the statement and the solution of mathematical optimization problems.
>
> In the past, each optimization problem had to be stated in the form of a standard model, for example, in a linear problem whether or not this was an accurate reflection of reality. This was necessary because only these standard models had theoretical solution procedures. If these procedures were followed, then a small amount of calculation produced the result.
>
> However, computers have become so fast and computer time so accessible and inexpensive that it is now possible to state any optimization problem (linear or nonlinear) as a completely accurate reflection of reality and let the computer search all the possible solutions (or a large sample of solutions) and produce the optimum regardless of the functional form of the problem.

Goodness of fit is often considered in conjunction with chi-square testing. Taylor [8] shows that it is not correct to use the chi-square test to test the goodness of fit of a linear compound graduation, and the Whittaker-Henderson method is a linear compound graduation. However, some recent literature (such as [1] and [5]) does not consider this result. For further discussion, see [4].

The multivariate calculus is a useful tool in many developments. Below are two examples.

(i) Assuming differentiability, one may rephrase the first part of Theorem 4.1 as

$$\frac{d}{d\lambda}[F(\boldsymbol{u}(\lambda)) + \lambda S(\boldsymbol{u}(\lambda))] \geq 0, \tag{4}$$

where $\boldsymbol{u}(\lambda) = \boldsymbol{u}^{\lambda}$ satisfies the equation

$$F'(\boldsymbol{u}(\lambda)) + S'(\boldsymbol{u}(\lambda)) = \mathbf{0}^T. \tag{5}$$

By the Chain Rule and the Product Rule,

$$\begin{aligned}
&\frac{d}{d\lambda}[F(\boldsymbol{u}(\lambda)) + \lambda S(\boldsymbol{u}(\lambda))] \\
&= F'(\boldsymbol{u}(\lambda))\boldsymbol{u}'(\lambda) + \lambda S'(\boldsymbol{u}(\lambda))\boldsymbol{u}'(\lambda) + S(\boldsymbol{u}(\lambda)) \\
&= \boldsymbol{0}^T\boldsymbol{u}'(\lambda) + S(\boldsymbol{u}(\lambda)) \qquad \because (5) \\
&= S(\boldsymbol{u}(\lambda)).
\end{aligned}$$

Thus we have (4).

(ii) The problem considered in [6] is the minimization of the quadratic form

$$q(\boldsymbol{u}) = (\boldsymbol{u}-\boldsymbol{u}'')^T W(\boldsymbol{u}-\boldsymbol{u}'') + (\boldsymbol{u}-\boldsymbol{s})^T V(\boldsymbol{u}-\boldsymbol{s}) + (K\boldsymbol{u})^T (K\boldsymbol{u}),$$

where W and V are diagonal matrices. The minimum vector $\boldsymbol{u}$ is the solution to the equation

$$q'(\boldsymbol{u}) = \mathbf{0}^T.$$

Since

$$q'(\boldsymbol{u}) = 2[(\boldsymbol{u}-\boldsymbol{u}'')^T W + (\boldsymbol{u}-\boldsymbol{s})^T V + \boldsymbol{u}^T K^T K],$$

the result of [6] follows.

REFERENCES

1. BENJAMIN, B., AND POLLARD, J. H. *The Analysis of Mortality and Other Actuarial Statistics*, London: Heinemann, 1980.
2. CONLEY, W. *Computer Optimization Techniques*, Princeton: Petrocelli Books, 1980.
3. GREVILLE, T. N. E. "A Fortran Program for Generalized Whittaker-Henderson Graduation," *ARCH*, 1982.1, Chicago: Society of Actuaries, 385–401.
4. HOEM, J. M. "A Contribution to the Statistical Theory of Linear Graduation," *Insurance: Mathematics and Economics*, III (1984), 1–17.
5. LONDON, R. L. *Graduation: The Revision of Estimates*, Part 5 Study Note 54-01-84, Itasca: Society of Actuaries, 1984.
6. LOWRIE, W. B. "An Extension of the Whittaker-Henderson Method of Graduation," *TSA*, XXXIV (1982), 329–50; Discussion 351–72.

7. SCHUETTE, D. R. "A Linear Programming Approach to Graduation," *TSA*, XXX (1978), 407–31; Discussion 433–45.
8. TAYLOR, G. C. "The Chi-Square Test of a Graduation by Linear-Compound Formula," *Bulletin de l'Association Royale des Actuaires Belges*, LXXI (1976), 26–42.

E.S. ROSENBLOOM:*

This paper presents algorithms for performing the Whittaker-Henderson graduation method of minimizing $F(\underset{\sim}{u}) + S(\underset{\sim}{u})$ over all $\underset{\sim}{u} = (u_1, u_2, \ldots, u_n)^T$. In section 2, the authors use the ℓ_p-norm with $1 < p < \infty$ to obtain the following nonlinear minimization problem:

$$\min_{\underset{\sim}{u}} [F(\underset{\sim}{u}) + \lambda S(\underset{\sim}{u})]$$

where

$$F(\underset{\sim}{u}) = \sum_{x=1}^{n} w_x \mid u_x - u_x'' \mid^p$$

and

$$S(\underset{\sim}{u}) = \sum_{x=1}^{n-z} \mid \nabla^z u_x \mid^p$$

($\lambda, p, w_1, w_2, \ldots, w_n, u_1'', u_2''$, are fixed constants).

The Newton-Raphson iterative technique is the algorithm used to find the optimal solution to this problem. In this algorithm, a sequence of vectors $\underset{\sim}{u}^k$ is generated using the recurrence relationship

$$\underset{\sim}{u}^{k+1} = \underset{\sim}{u}^k - [F''(\underset{\sim}{u}^k) + \lambda S''(\underset{\sim}{u}^k)]^{-1} [F'(\underset{\sim}{u}^k) + \lambda S'(\underset{\sim}{u}^k)]^T. \quad (1)$$

The strict convexity of the objective function ensures that if $\{\underset{\sim}{u}^k\}$ converges to $\underset{\sim}{u}$, then $\underset{\sim}{u}$ will be the unique optimal solution to the nonlinear program.

The Newton-Raphson algorithm is one of the oldest numerical techniques available for minimizing nonlinear functions. It has the advantage that when it converges, it converges at a quadratic rate. However, the Newton-Raphson method is rarely used today to solve nonlinear problems because it has certain drawbacks.

Formula (1) does not ensure a decrease in the function value at each iteration. In other words, it is possible that

$$F(\underset{\sim}{u}^{k+1}) + \lambda S(\underset{\sim}{u}^{k+1}) > F(\underset{\sim}{u}^k) + \lambda S(\underset{\sim}{u}^k).$$

To remedy this situation the modified Newton-Raphson formula is often used:

*Dr. Rosenbloom, not a member of the Society, is an Assistant Professor with the Department of Actuarial and Management Sciences, University of Manitoba.

$$\underset{\sim}{u}^{k+1} = \underset{\sim}{u}^{k} - \theta^{k} [F'' (\underset{\sim}{u}^{k}) + \lambda S'' (\underset{\sim}{u}^{k})]^{-1} [F' (\underset{\sim}{u}^{k}) + \lambda S' (\underset{\sim}{u}^{k})]^{T} \quad (2)$$

where θ^k is a scalar chosen so that

$$F (\underset{\sim}{u}^{k+1}) + \lambda S (\underset{\sim}{u}^{k+1}) < F (\underset{\sim}{u}^{k}) + \lambda S (\underset{\sim}{u}^{k}).$$

In one variation, θ^k is chosen to minimize $F (\underset{\sim}{u}^{k} + \theta \underset{\sim}{d}^{k}) + \lambda S (\underset{\sim}{u}^{k} + \theta \underset{\sim}{d}^{k})$ with respect to θ where $\underset{\sim}{d}^{k}$ is the search direction

$$- [F'' (\underset{\sim}{u}^{k}) + \lambda S'' (\underset{\sim}{u}^{k})]^{-1} [F' (\underset{\sim}{u}^{k}) + \lambda S' (\underset{\sim}{u}^{k})]^{T}.$$

Another drawback of the Newton-Raphson method is that even with the assumption of strict convexity of the objective function $F (\underset{\sim}{u}) + \lambda S (\underset{\sim}{u})$, the Hessian matrix $F'' (\underset{\sim}{u}^{k}) + \lambda S'' (\underset{\sim}{u}^{k})$ may be singular. In that case, formula (1) would be undefined.

For large values of n, the most serious drawback of the Newton-Raphson method is the enormous amount of computation required at each iteration. Even exploiting the fact that the Hessian matrix $F'' (\underset{\sim}{u}^{k}) + \lambda S'' (\underset{\sim}{u}^{k})$ is symmetric, formula (1) requires computing $n^2 + n$ second partial derivatives and $2n$ first partial derivatives. In addition, a system of equations needs to be solved in order to obtain the search direction $\underset{\sim}{d}^{k}$.

To avoid the drawbacks of the Newton-Raphson method a number of techniques have been developed over the last twenty-five years. The most popular of these techniques are the Quasi-Newton methods. The Quasi-Newton methods generate a sequence of vectors $\{\underset{\sim}{u}^{k}\}$ by a recurrence relationship of the form.

$$\underset{\sim}{u}^{k+1} = \underset{\sim}{u}^{k} - \theta^{k} H^{k} [F' (\underset{\sim}{u}^{k}) + \lambda S' (\underset{\sim}{u}^{k})]^{T}. \quad (3)$$

H^k is an $n \times n$ matrix which may approximate the inverse Hessian matrix $[F'' (\underset{\sim}{u}^{k}) + \lambda S'' (\underset{\sim}{u}^{k})]^{-1}$. θ^k is chosen to minimize $F (\underset{\sim}{u}^{k} + \theta \underset{\sim}{d}^{k}) + \lambda S (\underset{\sim}{u}^{k} + \theta \underset{\sim}{d}^{k})$ with respect to θ with $\underset{\sim}{d}^{k}$ being the direction $- H^k [F' (\underset{\sim}{u}^{k}) + \lambda S' (\underset{\sim}{u}^{k})]^{T}$.

In general, the Quasi-Newton methods require considerably less computation than the Newton-Raphson methods. They do not require the computation of second partial derivatives. In addition, they tend to be more robust than the Newton-Raphson methods.

The various Quasi-Newton methods differ in how the matrix H^k is obtained. Numerical experiments have indicated that the most successful of the Quasi-Newton methods is the Broyden-Fletcher-Goldfarb-Shanno algorithm. In this algorithm, the matrix H^k is obtained using the formula

$$H^{k+1} = H^{k} + \left(\frac{1 + (\underset{\sim}{q}^{k})^{T} H^{k} \underset{\sim}{q}^{k}}{(\underset{\sim}{q}^{k})^{T} \underset{\sim}{p}^{k}} \right) \frac{\underset{\sim}{p}^{k} (\underset{\sim}{p}^{k})^{T}}{(\underset{\sim}{p}^{k})^{T} \underset{\sim}{q}^{k}} - \frac{\underset{\sim}{p}^{k} (\underset{\sim}{q}^{k})^{T} H^{k} + H^{k} \underset{\sim}{q}^{k} (\underset{\sim}{p}^{k})^{T}}{(\underset{\sim}{q}^{k})^{T} \underset{\sim}{p}^{k}} \quad (4)$$

where

H^o is any positive definite matrix,

$\underset{\sim}{p}^k = \underset{\sim}{u}^k - \underset{\sim}{u}^{k-1}$, and

$$\underset{\sim}{q}^k = [\,F'(\underset{\sim}{u}^k) + \lambda\, S'(\underset{\sim}{u}^k)\,]^T - [\,F'(\underset{\sim}{u}^{k-1}) + \lambda\, S'(\underset{\sim}{u}^{k-1})\,]^T.$$

A more complete discussion of Quasi-Newton methods and other alternatives to Newton-Raphson can be found in [4] or [5].

REFERENCES

1. AVRIEL, M. *Nonlinear Programming*, Englewood Cliffs: Prentice Hall, 1976.
2. FLETCHER, R. "A New Approach to Variable Metric Algorithms," *Computer Journal* 13 (1970), 317–22.
3. GILL, P.E., AND MURRAY, W. "Quasi-Newton Methods for Unconstrained Optimization," *Journal of the Institute of Mathematics and ITS Applications* 9 (1972), 91–108.
4. GILL, P.E., MURRAY, W., AND WRIGHT, M.H. *Practical Optimization*, New York: Academic Press, 1981.
5. LUENBERGER, D.G. *Linear and Non-linear Programming*, Reading, Mass.: Addison-Wesley, 1984.
6. MURRAY, W. *Numerical Methods for Unconstrained Optimization*, New York: Academic Press, 1972.

(AUTHORS' REVIEW OF DISCUSSION)

FUNG YEE CHAN, LAI K. CHAN, AND MAN HEI YU:

Dr. Shiu generalizes $F + \lambda S$ using a formulation which covers a wide range of cases. Multivariate calculus can then be applied to the graduated values $\underset{\sim}{u}^{\lambda}$ and used to derive the monotone properties.

The problem of minimizing

$$\sum_x w_x|u_x - u_x''| + \lambda \max_y |\Delta^z u_y|$$

and other related problems, of which the norms of F and S are different and are ℓ_1, ℓ_2 or ℓ_∞, have been investigated by us in a separate study.

We agree with Dr. Shiu's comment that, due to the advances of computer technology, new approaches in the formulation and solution procedures for the Whittaker-Henderson graduation should be explored. Recently, we have been working on a statistical data analysis approach of selecting λ.

Dr. Rosenbloom gives a comprehensive description of contemporary techniques which improve the traditional Newton-Raphson method.

In his 1974 Part 5 Society of Actuaries Study Note on graduation, Dr. Thomas Greville elegantly used linear algebra to formulate and solve the

Whittaker-Henderson graduation problem. His work has inspired using modern mathematics in research work on graduation. This paper and the two discussions represent some of the inspiration.

We would like to take this opportunity to thank Dr. Greville, who has contributed so much to the development of graduation methodology.