ACTUARIAL RESEARCH CLEARING HOUSE
1994 VOL. 1

# Bivariate Schuette Graduation

Esther Portnoy
University of Illinois at Urbana–Champaign

*Abstract*

Two–dimensional data can be smoothed by minimizing a combination of the sum of weighted absolute deviations from initial estimates and the sums of absolute values of differences in the two directions. The techniques of parametric programming make it feasible to explore the different results obtained by varying the emphasis given to fit *vs* smoothness.

## 1. *Introduction*

In 1978 Donald Schuette suggested a variation in the familiar Whittaker–Henderson method of graduation, replacing squares by absolute values in the measures of both fit and smoothness and finding the solution via linear programming. Because the method is described very briefly in [London, 1985], which is the official syllabus of SOA Course 165, many actuaries will recognize the name; but probably few have studied it carefully. Standing in the way of wider practical use of Schuette's method are a lack of appreciation for the rationale and (possibly more important) perceived computational difficulties. Recent work in parametric programming has made it possible to streamline the computations, so that one can solve completely the Schuette problem (that is, generate *all* the critical values of the control parameter as well as the vectors of graduated values) as quickly as one can generate solutions to the Whittaker problem for a few different parameter values.

Generalization to bivariate graduation initially appears to be straightforward, but some interesting complications arise. Given two–dimensional initial estimates $u_{ij}$, we seek $v_{ij}$ to minimize

$$M(\alpha,\beta) = \Sigma\, w_{ij}\,|v_{ij} - u_{ij}| \;+\; \alpha\,\Sigma\,|\Delta_{\downarrow}^{z}\, v_{ij}| \;+\; \beta\,\Sigma\,|\Delta_{\rightarrow}^{\zeta}\, v_{ij}|\ ,$$

where $\Delta_{\downarrow}$ represents difference in the "vertical" direction (for example, $\Delta_{\downarrow} v_{11} = v_{21} - v_{11}$) and $\Delta_{\rightarrow}$ represents difference in the "horizontal" direction ($\Delta_{\rightarrow} v_{11} = v_{12} - v_{11}$). In practice we rarely begin with specific values for $\alpha$ and $\beta$; rather, we use the form above as an aid in searching for a good graduation. Since the minimization can be viewed as a linear programming problem, the optimal solution for any particular $(\alpha,\beta)$ will occur at one or more "basic" solutions, of which there are only finitely many, and the positive quadrant in the $(\alpha,\beta)$ plane is divided into regions in which the optimal graduation is constant. Thus we shift the emphasis of the search: instead of beginning with $(\alpha,\beta)$ and looking for optimal $v_{ij}$, we begin with $v_{ij}$ and ask for what set of $(\alpha,\beta)$ the indicated graduation is optimal. This subtle change turns out to be remarkably fruitful.

The choice of absolute values in the measure M has been discussed at length. The strongest competitor, of course, uses squared deviations for the terms relating to fit, and squares of differences in both directions — two–dimensional Whittaker graduation. Compared to this $l_2$–measure, the $l_1$–measure used here is more robust: if some of the $u_{ij}$ have values out of line with others, then measuring fit by absolute deviation rather than by squared deviation will result in a graduation that is less influenced by those outliers. Specifically, if one of the $u_{ij}$ is changed but does not cross the graduation curve (so that the sign of $(v_{ij} - u_{ij})$ is not changed), then the graduated values do not change at all.

There are other candidates for robust measures of fit (see, for example, Klugman's discussion of [Schuette, 1978]), but they seem to present considerable computational difficulties. A final advantage of the $l_1$ method (which will not be demonstrated here, but see [Koenker and Bassett, 1978]) is the prospect of providing estimates of the likely spread of values. A single graduation by the $l_1$ method corresponds in a natural way to a median estimate; by changing the fit measure we can obtain, say, 25th and 75th percentiles.

## 2. *Mechanics of the graduation*

In this section we present the parametric–programming method of determining the region in which a particular solution is optimal, and of proceeding from one optimal solution to an adjacent one. We will ignore, for the moment, certain complications that can arise. It may be helpful to follow the very simple example presented in Exhibit 1 as the needed quantities are defined. A more general discussion of parametric programming can be found in [Guddat *et al.*, 1985].

We begin with initial estimates $u_{ij}$ $(1 \le i \le m,\ 1 \le j \le k)$. Set $mk = n$, the number of "coordinate" variables, and $N = n + k(m-z) + m(k-\zeta)$. Define an $N \times n$ design matrix D made up of the $n \times n$ identity matrix, followed by the differencing matrices $\mathbf{k}_{\downarrow z}$ and $\mathbf{k}_{\rightarrow \zeta}$. At a "basic" solution to the problem (in the linear–programming sense) some observed values are matched, some differences vanish, and the remaining quantities can be expressed in terms of the basic variables. Let $D_1$ be the nonsingular submatrix made up of those rows of D corresponding to the basic variables, and $D_2$ the $(N-n) \times n$ matrix of remaining rows. Define a column vector **b** whose elements are, first, the weights $w_{ij}$, then $\alpha$ for each of the $k(m-z)$ vertical differences, and finally $\beta$ for each of the $m(k-\zeta)$ horizontal differences. Divide **b** into vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ just as D was divided. Let $D_2^h$ be the matrix obtained by multiplying each row of $D_2$ by the appropriate factor among

$$w_{ij} \operatorname{sgn} (v_{ij} - u_{ij})$$

$$\alpha \operatorname{sgn} (\Delta_{\downarrow}^{z} v_{ij})$$

and

$$\beta \operatorname{sgn} (\Delta_{\rightarrow}^{\zeta} v_{ij}) .$$

(We assume for now that none of the terms above vanish. Degeneracy is a problem that will be treated in Section 3.) Finally, let **y** be the row vector obtained by summing the columns of $D_2^h$. The elements of **y** give those terms of M that do not vanish near the basic solution, each with appropriate sign, as functions of the coordinate variables $v_{ij}$.

Right–multiplication by $D_1^{-1}$ gives the same terms but now as functions of the basic variables.

If one of the basic variables increases (resp. decreases) while the others are unchanged, the resulting change in M is proportional to the appropriate element of $\mathbf{b}_1$ ($w_{..}$ or $\alpha$ or $\beta$), plus (resp. minus) the corresponding element of $\mathbf{y}\cdot D_1^{-1}$. For a particular $(\alpha,\beta)$, M will be minimized at the basic solution in question provided *all* of these changes are non–negative; that is, coordinate by coordinate,

$$\mathbf{b}_1^{\mathsf{T}} \pm \mathbf{y}\cdot D_1^{-1} \geq 0 .$$

We have thus a set of $2n$ linear inequalities in $(\alpha,\beta)$, together with the constraints $\alpha$, $\beta \geq 0$. The region in which these are satisfied simultaneously is either the empty set or a convex polygonal region in the $(\alpha,\beta)$ plane. If the number of variables is large, determining the region may be difficult. Some suggestions are made in Section 4.

Next, define another column vector $\mathbf{m}$ of current values (ie $v_{ij} - u_{ij}$ or $\Delta_{\downarrow}^{z} v_{ij}$ or $\Delta_{\rightarrow}^{\zeta} v_{ij}$), which must therefore have 0's in the "basic" rows. Note that all the information about a particular basic solution is conveyed by the matrix $[D\cdot D_1^{-1}:\mathrm{m}:\mathrm{b}]$. Imbedded in $D\cdot D_1^{-1}$ is an identity matrix, marking the basic variables; the remaining rows $D_2\cdot D_1^{-1}$ give the non–basic variables in terms of the basic ones; and we can develop the vector $\mathbf{y}\cdot D_1^{-1}$ by multiplying the $j$th row of $D\cdot D_1^{-1}$ by $b_j \operatorname{sgn}(m_j)$ — with the understanding that $\operatorname{sgn}(0) = 0$ — and then adding each column.

In particular, if $D_1 = I_{n\times n}$ then:

$$\mathbf{m} = [0,\dots,0,\Delta_{\downarrow}^{z} u_{11},\dots,\Delta_{\rightarrow}^{\zeta} u_{m:k-\zeta}]^{\mathsf{T}}$$

$$v_{ij} \equiv u_{ij} \qquad \text{(the no–graduation case)}$$

and

$$\mathbf{b}_1 = [w_{11},\dots,w_{mk}]^{\mathsf{T}}$$

The resulting inequalities are all of the form $-w_{ij} \leq a_{ij}\,\alpha + b_{ij}\,\beta \leq w_{ij}$, where the $a_{ij}$ and $b_{ij}$ are algebraic sums of the binomial coefficients of orders $z$ and $\zeta$ respectively. It is obvious that (0,0) satisfies each of these inequalities, as do $(\alpha_0, 0)$ and $(0,\beta_0)$, where $\alpha_0 = \min\{w_{ij}/|a_{ij}|\} > 0$ and $\beta_0 = \min\{w_{ij}/|b_{ij}|\} > 0$. Thus the no–graduation solution is optimal in a region which includes a nontrivial neighborhood of (0,0). In the one–dimensional case, Schuette gave an explicit formula for the first critical value of the smoothing parameter, below which the optimal solution is no–graduation. To date we cannot give an analogous formula in the two–dimensional case. It can be shown that under some modest assumptions the region is bounded; it is not necessarily a triangle.

Transition from one basic solution to an adjacent one is effected by a single simplex pivot, as follows. In the process of determining the region in which a particular basic solution is optimal, we must associate with each actual edge of the region one of the $2n$ inequalities. This singles out a particular variable to leave the basic set, *and* tells us whether to increase it or decrease it in order to cross the edge. Suppose that a certain edge is defined by the constraint corresponding to *increasing* the $r$th basic variable $x_r$ (which may be some coordinate variable $v_{ij}$ or some vertical or horizontal difference). The row of $D \cdot D_1^{-1}$ corresponding to $x_r$ has all 0's except 1 in the $r$th column. The other elements in the $r$th column tell us how the other variables will change as we change $x_r$ but hold the other basic variables fixed. Some of the non–basic variables may increase, others decrease, and some may remain unchanged. (All the other *basic* variables have 0's in the $r$th column of $D \cdot D_1^{-1}$.) We need to increase $x_r$ until some currently non–basic variable reaches its basic value.

Suppose a particular nonbasic variable $y_s$ has a positive entry in the $r$th column of $D \cdot D_1^{-1}$ and a positive entry in **m**. That is, $y_s$ is already above its basic value and will only increase further as $x_r$ is increased. We need to find a nonbasic variable such that the ratio of the $m$–coordinate (ie the current value) to the entry in the $r$th column of $D \cdot D_1^{-1}$ is negative, and if there are several, choose the one with the smallest absolute value to define the entering basic variable. If the constraint had involved *decreasing* $x_r$ then we would look for a minimum *positive* ratio $m_{\cdot}/(D \cdot D_1^{-1})_{\cdot r}$. (A tie will lead to degeneracy at the new basic solution.) Once the entering variable is selected, we do a simplex pivot, using elementary column operations so that the row of the new $[D \cdot D_1^{-1} : \mathbf{m}]$ corresponding to the entering variable has all 0's except a 1 in the $r$th column. (The vector **b** does not change.) The only difference between this and the standard simplex method is that we must determine the appropriate sign for the ratio at each pivot.

## 3. *Degeneracy*

It can happen at a basic solution that some of the non–basic variables take on their basic values; that is, that M has more than $n$ vanishing terms at that solution. The terms corresponding to non–basic variables that happen to vanish will increase (if they change at all) with both increases and decreases in the basic variables, instead of increasing with some changes and decreasing with others. We can alter the inequalities to reflect this, but the altered set will not necessarily suffice for optimality. We would need to generate every possible basis among the variables that take on their basic values, generate $2n$ inequalities for each basis, and combine them (some inequalities will appear in several sets) to form a defining set of inequalities.

In the one–dimensional case degeneracy is a mere nuisance: it arises when by happenstance some of the initial values lie on a small–degree polynomial curve, and it can be avoided by "dithering" the initial estimates (adding or subtracting small amounts that will be lost in round–off). In the two–dimensional case the problem is more severe, since as originally stated there are *systemic* dependencies among the variables. For example, when both $\alpha$ and $\beta$ are sufficiently large the optimal solution has all differences vanishing, and a few of the initial estimates matched. Exhibit 1 is atypical in exhibiting degeneracy only at this "last" stage, and having only one extra 0 in the final m vector. In what is still an impractically small example, $m = k = 5$, $z = \zeta = 2$, the last stage has at least 33 variables at their basic values (all 30 differences vanishing, and three of the initial estimates determining a plane), of which we require only 25 for a basis. Moreover, there will be degeneracies "earlier" as the vanishing of certain sets of differences will force the vanishing of others.

Fortunately, there is an easy way around this difficulty. Instead of dithering the initial estimates $u_{ij}$, we can dither the design matrix, by adding to each initially nonzero entry of the original $D_2$ submatrix a random quantity $\epsilon$, say, uniformly distributed on $(-10^{-5},+10^{-5})$. (Remember that the entries in the initial design matrix are integers). What that means is that we will be minimizing not the original M but a slightly different function. In effect, we "trick" the program into running through different bases for the same solution. This is illustrated for Exhibit 1. For reasonable–sized examples, some investigation will be required, to make sure that the small changes introduced do not accumulate to significance. It may be reasonable to do the longer analysis for a few solutions among which one finally chooses the desired graduation.

## 4. *Determining the region, when the number of variables is large*

In Exhibit 1 it was possible to write down the eight inequalities generated at each stage and reduce them by hand to a minimal set, one for each edge of the region. If the number of variables is large we need to automate this process. In addition, we need to associate with each edge the basic variable that will change (and in which direction) as we cross that edge, in order to be able to go easily to an adjacent solution. A method that seems promising is to reformulate our problem into a one–dimensional parametric programming problem. We illustrate this at the "initial", no–graduation solution. It can be applied beginning with any basic solution, though it does help to start at one known edge.

We ask, in what region (that is, for what values of $\alpha$ and $\beta$) is the no–graduation solution optimal? We know that it is optimal at (0,0). Moving along the $\alpha$–axis, we formulate the one–dimensional parametric programming problem: find the interval in which the no–graduation solution minimizes $M(\alpha,0)$. The mechanics described in Section 2 apply perfectly well; we simply arrive at a system of inequalities $-w_{ij} \leq a_{ij}\alpha \leq w_{ij}$, which are clearly satisfied as long, and only as long, as $\alpha \leq \min\{w_{ij}/|a_{ij}|\} = \alpha_0$. Even if $n$ is large, it is not hard to pick out the smallest among a set of $n$ numbers, and to keep track of the indices $(i,j)$ that give rise to it (assuming it is unique).

So we know that one edge of the region of optimality is (0,0), $(\alpha_0,0)$, *and* we know that the adjacent edge lies along the line $a_{ij}\alpha + b_{ij}\beta = w_{ij}\,\mathrm{sgn}(a_{ij})$. We parametrize this line as $(\alpha_0 - t\cdot b_{ij}/a_{ij},\ t)$, and ask for what interval of (non–negative) $t$ values the no–graduation solution is optimal for $M(\alpha_0 - t\cdot b_{ij}/a_{ij},\ t)$. Again, viewing this as a one–dimensional parametric programming problem leads us easily to a maximal $t_1$ where a particular inequality becomes an equality. $(\alpha_0 - t_1 b_{ij}/a_{ij},\ t_1)$ is a vertex of the region of optimality, and we move along the next segment as far as possible. Repeating this process,we eventually come to a vertex on the $\beta$–axis, which can only be $(0,\beta_0)$; moving down the axis to (0,0) completes our circuit of the region of optimality.

We do not yet have a program running to accomplish this part of the solution. Clearly, considerable care will be needed to prevent missing the completion of a circuit because of round–off error. Some "subroutines" that have been used on somewhat more complicated examples than the one in Exhibit 1 should form the foundation of a more complete program.

## 5. *Moving toward a good graduation*

For examples of even moderate size, it may not be practical to find the complete set of optimal solutions. The natural goal, of course, is to find (and justify) a good graduation, so we may be content to find a collection of solutions among which we can reasonably expect to find one that is satisfactory. This section presents one method by which this could be done.

Suppose we have an *a priori* estimate for suitable values of $\alpha$ and $\beta$, or at least an initial estimate of the ratio $\alpha/\beta$. Carry out a one–dimensional routine generating the critical values and optimal solutions for $M(t\alpha,t\beta)$ , $0 < t < \infty$ . Note a few good candidates, and find the regions of optimality for each. If nothing really satisfactory is found — or if the regions seem to be very small — one can proceed along one or more lines transverse to the first one.

For truly one–dimensional problems, there exists some quite fast software that not only generates the critical values of the parameter, but graphs the graduated values at each stage. The graduator can watch as the process develops, and stop when a suitable balance between fit and smoothness is attained. (In fact, for this reason Koenker suggests that it may generally be better to start with one of the existing good algorithms to obtain the $l_1$–fit polynomial (all differences vanishing), then proceed downward by parametric programming through smaller values of the smoothness parameter. Generally one will stop well before the no–graduation limit is reached.) Analogous software displaying two–dimensional solutions would certainly facilitate the hunt for a suitable graduation, but may be beyond our currect capabilities.

### *References*


Guddat, J., F. Guerra Vasquez, K. Tammer and K. Wendler (1985). *Multiobjective and Stochastic Optimization Based on Parametric Optimization*, Akademie–Verlag (Berlin).

Koenker, R., and G. Bassett (1978). "Regression Quantiles," *Econometrica*, 46, 33–50.

London, D. (1985). *Graduation: the Revision of Estimates*, ACTEX.

Schuette, D.R. (1978). "A Linear Programming Approach to Graduation," *Trans. Soc. of Actuaries*, 30, 407–445 (including discussion).

# Exhibit 1

Let $\mathbf{u} = \begin{bmatrix} 1 & 3 \\ 5 & 2 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 & 5 \\ 7 & 3 \end{bmatrix}$. Of necessity, $\zeta = z = 1$, so $n = 4$, $\mathrm{N} = 8$.

The design matrix $\mathrm{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$. Consider the basic solution $\mathbf{v} = \begin{bmatrix} 3 & 3 \\ 5 & 3 \end{bmatrix}$.

The basic variables are $v_{12}$, $v_{21}$, $\Delta_{\downarrow} v_{12}$ and $\Delta_{\rightarrow} v_{11}$. Near this point we have $\mathrm{M} = 5|v_{12}-3| + 7|v_{21}-5| + \alpha|\Delta_{\downarrow} v_{12}| + \beta|\Delta_{\rightarrow} v_{11}| + 2(v_{12}-\Delta_{\rightarrow} v_{11}-1) + 3(v_{12}+\Delta_{\downarrow} v_{12}-2) + \alpha[v_{21}-v_{12}+\Delta_{\rightarrow} v_{11}] - \beta[v_{12}-v_{21}+\Delta_{\downarrow} v_{12}]$. Next,

$$[\mathrm{D}_1{:}\mathrm{b}_1] = \begin{bmatrix} 0 & 1 & 0 & 0 & : 5 \\ 0 & 0 & 1 & 0 & : 7 \\ 0 & -1 & 0 & 1 & : \alpha \\ -1 & 1 & 0 & 0 & : \beta \end{bmatrix}, \quad [\mathrm{D}_2^h{:}\mathrm{b}_2] = \begin{bmatrix} 2 & 0 & 0 & 0 & : 2 \\ 0 & 0 & 0 & 3 & : 3 \\ -\alpha & 0 & \alpha & 0 & : \alpha \\ 0 & 0 & \beta & -\beta & : \beta \end{bmatrix}$$

$$\mathrm{y} = [\, 2-\alpha,\ 0,\ \alpha+\beta,\ 3-\beta \,]$$

$$\mathrm{y} \cdot \mathrm{D}_1^{-1} = [5-\alpha-\beta, \alpha+\beta, 3-\beta, \alpha-2]$$

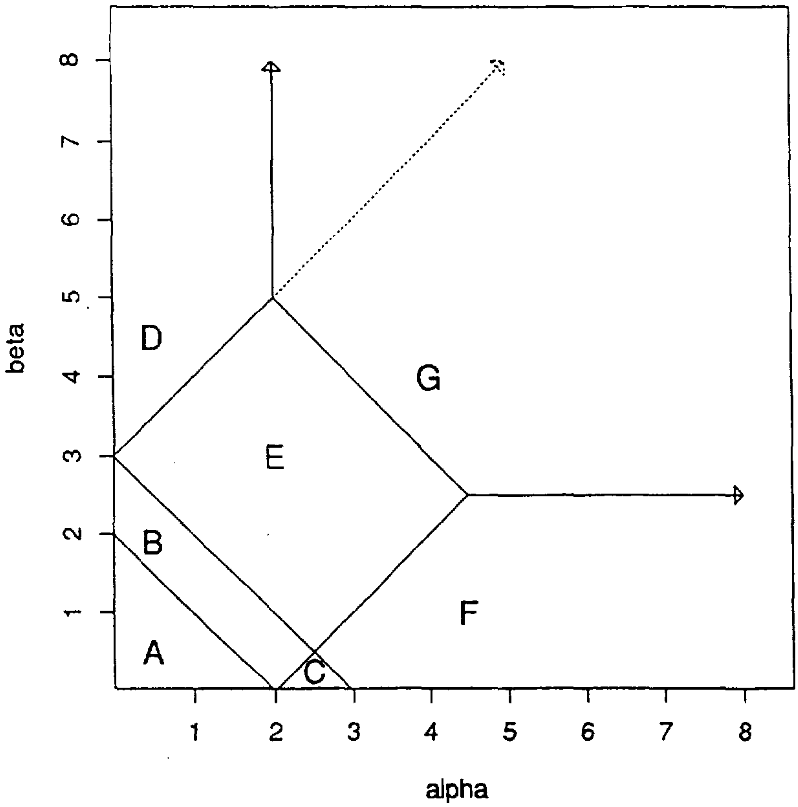
The basic solution is optimal provided $\mathrm{b}_1^{\mathrm{T}} \pm \mathrm{y} \cdot \mathrm{D}_1^{-1} \geq 0$ ie

$$5 \pm (5 - \alpha - \beta) \geq 0$$
$$7 \pm (\alpha + \beta) \geq 0$$
$$\alpha \pm (3 - \beta) \geq 0$$
$$\beta \pm (\alpha - 2) \geq 0$$

$$[\mathrm{D} \cdot \mathrm{D}_1^{-1}{:}\mathrm{m}{:}\mathrm{b}] = \begin{bmatrix} 1 & 0 & 0 & -1 & : 2 & : 2 \\ 1 & 0 & 0 & 0 & : 0 & : 5 \\ 0 & 1 & 0 & 0 & : 0 & : 7 \\ 1 & 0 & 1 & 0 & : 1 & : 3 \\ -1 & 1 & 0 & \textcircled{1} & : 2 & : \alpha \\ 0 & 0 & 1 & 0 & : 0 & : \alpha \\ 0 & 0 & 0 & 1 & : 0 & : \beta \\ 1 & -1 & 1 & 0 & : -2 & : \beta \end{bmatrix} \underset{\rightarrow}{\text{pivot}} \begin{bmatrix} 0 & 1 & 0 & -1 & : 4 & : 2 \\ 1 & 0 & 0 & 0 & : 0 & : 5 \\ 0 & 1 & 0 & 0 & : 0 & : 7 \\ 1 & 0 & 1 & 0 & : 1 & : 3 \\ 0 & 0 & 0 & 1 & : 0 & : \alpha \\ 0 & 0 & 1 & 0 & : 0 & : \alpha \\ 1 & -1 & 0 & 1 & : -2 & : \beta \\ 1 & -1 & 1 & 0 & : -2 & : \beta \end{bmatrix}.$$

The complete solution divides the positive $(\alpha,\beta)$ quadrant into the seven regions shown in Figure 1.

FIGURE 1. REGIONS OF OPTIMALITY




The "graduated" values in the several regions are:

A: $\begin{bmatrix}1 & 3\\ 5 & 2\end{bmatrix}$ B: $\begin{bmatrix}3 & 3\\ 5 & 2\end{bmatrix}$ C: $\begin{bmatrix}5 & 3\\ 5 & 2\end{bmatrix}$ D: $\begin{bmatrix}3 & 3\\ 5 & 5\end{bmatrix}$

E: $\begin{bmatrix}3 & 3\\ 5 & 3\end{bmatrix}$ F: $\begin{bmatrix}5 & 3\\ 5 & 3\end{bmatrix}$ G: $\begin{bmatrix}3 & 3\\ 3 & 3\end{bmatrix}$

"Dithering" the design matrix:

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -.99997 & 0 & .99995 & 0 \\ 0 & -1.00009 & 0 & 1.00008 \\ -.99995 & .99999 & 0 & 0 \\ 0 & 0 & -.99998 & .99991 \end{bmatrix}$$

splits region G into two parts along the dotted line, and causes insignificant changes in the graduated values and regions of optimality. For example, the vertices of the regions of optimality, under dithering as above, are: (2.00006,0), (2.99976,0), (2.49995,0.49990), (0,2.00010), (4.50014,2.50013), (0,3.00027), (1.99980,5.00041), to five decimal places; these agree with the original vertices to three decimal places. There are similar fourth– and fifth–place differences in the "graduated" values.