XIV CONGRESS OF THE INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY HAMBURG 1980
Commission III Presented paper

E. Wild, Stuttgart University

INTERPOLATION WITH WEIGHT-FUNCTIONS - A GENERAL INTERPOLATION METHOD

SUMMARY

Analysis of existing methods shows that interpolation can be described by weight functions which decompose an interpolation surface in components which are related to the reference points. They trace the effects of variation of reference values onto the interpolation. Alternately a rather general interpolation method can be developed based on weight functions. By specifying certain features the method is adapted to various applications, in particular to high accuracy DTM interpolation.

Interpolation is generally used for the approximate description of functional relationships which are not explicitely known but of which a limited number of functional values is given. In addition we demand from the interpolation specific features, according to our knowledge or expectation about the relationship. The problem has no unique solution, therefore a number of interpolation methods have been developed, based on very different principles.

The interpolation of digital terrain models (DTM) is a demonstrative example. In an invited paper /6/ for the ISP congress 1976 in Helsinki Shut has given a review about the interpolation methods used for DTM interpolation. One can distinguish polynomial interpolation (Bosman, Eckhart, Kubik /2/), least squares interpolation (Kraus /5/), summation of surfaces (Hardy /4/), and interpolation with finite elements (Ebner, Reiss /3/).

In this paper the interpolation with weight functions is presented as a method for analysing known interpolation methods and as a new interpolation method. By using weight functions the inherent relation between the reference points and the interpolation surface becomes evident. Thus it is possible to show the specific features of a known interpolation method. It is also possible, however, to create weight functions with certain features, which are related to specific properties of the resulting interpolation. Therefore the interpolation with weight functions can be considered a general interpolation method, which can be adapted to special interpolation problems.

It is convenient to create weight functions as linear combinations of certain base functions. All required types of weight functions can be obtained in this way. The method of base functions has been applied for interpolation with weight functions for the purpose of high accuracy DTM. The important features of the method are shown in this paper. The interpolation method was realized with a computer program related to the DTM and contour interpolation program SCOP (see Kraus /5/, Stanger /7/, Assmus /1/). The least squares interpolation of the SCOP program was replaced by interpolation with weight functions. Three examples of derived contour lines are presented in fig. 3.

1. The principle of weight functions

DTM interpolation is an interpolation problem with a 2-dimensional distribution of the reference points. The terrain surface is to be described by an interpolation surface. A functional definition of the terrain surface is not known, but we have a description by reference points. In addition we have some knowledge about the features of the terrain surface and we impose the same or similar features on the interpolation surface. For various applications and specifications the feature of the surface can be chosen differently.

Most interpolation methods are based on a linear relation between the interpolation value z_i and the reference values r. It means that the interpolation values z_i of the interpolation surface are described as the weighted sum of the reference values r:

$$z_i = w_1 \cdot r_1 + w_2 \cdot r_2 + \ldots w_n \cdot r_n \qquad (1)$$

Accordingly, by the interpolation with weight functions, each interpolation value is decomposed in components, which can be related to the reference points. The weighted contribution of each reference value varies at each interpolation point. Therefore the weight of any reference point j is regarded a function of the coordinates x_i und y_i of the interpolation point. For any reference point j with the reference value r_j we get the weight function, for the interpolation at point i, as

$$w_j = f_j \ (x_i, y_i) \qquad (2)$$

The weight function w_j of the reference point j defines the weight with which the reference value r_j contributes to the computation of the interpolation value at each interpolation point. The weight function describes the effect of a reference point on the interpolation surface. Such weight functions will be used now for analysing known interpolation methods and for setting up a new interpolation method.

2. The deduction of weight functions of known interpolation methods

In general the interpolation methods describe the interpolation surface as a function with the coordinates x_i and y_i of the interpolation point as parameters. This description can be interpreted as an interpolation with weight functions, of which the weight functions can be established. The deduction of weight functions can be done for the interpolation with polynomials, for the least squares interpolation and for the interpolation with the summation of surfaces. For the interpolation with finite elements, the weight functions can be described by grid values, in the same way as the method presents the interpolation surface.

In this chapter the general deduction of weight functions will be shown and as an example the weight functions of the least squares interpolation will be regarded in detail. Different characteristics appear, when the reference points have a regular or an irregular distribution.

2.1 The general case

In general an interpolation surface is defined as a function of the coordinates u and v of the reference points, the reference values r and the coordinates x_i and y_i of the interpolation point, which are used as parameters for the description of a surface

$$z_i = f\,(x_i , y_i \; ; u_1 \ldots u_n , v_1 \ldots v_n , r_1 \ldots r_n) \qquad (3)$$

The previously mentioned four interpolation methods can be described by the formula

$$z_i = \underline{f}^T \cdot \underline{B} \cdot \underline{r} \qquad (4)$$

where

$\underline{r}$ is the vector of the n reference values
$\underline{B}$ is a m·n matrix, where the elements are functions of the coordinates u, v of the reference points
$\underline{f}$ is a vector with m elements, which contain the parameters x_i and y_i for the surface discription

For the interpolation with finite elements, the vector $\underline{f}$ is replaced by a l·m matrix $\underline{F}$, and the left side of the equation becomes a vector $\underline{z}$, which contains the interpolated grid values. If we compare formula (4) with the interpolation with weight functions (1), we can deduce the weight functions as

$$\underline{w} = \underline{f}^T \cdot \underline{B} \qquad (5)$$

where the vector $\underline{w}$ contains the n weight functions. We get the weight function w_j which is related to the reference point j, as

$$w_j = \underline{f}^T \cdot \underline{B} \cdot \underline{e}_j \qquad (6)$$

where the vector $\underline{e}_j$ contains what we call standard values of the weight-functions w_j.

The standard values define values for the weight functions in the reference points. Each weight function is related to a certain reference point and its standard value in this point is 1. All other standard values of a weight function are 0. The standard values have their characteristic importance by the fact, that if all weight functions pass through their respective standard values, the interpolation surface will exactly pass through all reference values.

2.2 The weight functions of least squares interpolation

The least squares interpolation is based on statistical considerations. By analysing the reference points with regard to correlation the statistical behaviour of the surface can be described by a covariance function. As covariance function usually the function

$$Cov = (1-f)e^{-s^2/m^2}$$

is used, which refers to the variance 1. After the variances and the covariance-function are known, the interpolation surface is described by the formula

$$z_i = \underline{f}^T \underline{C}^{-1} \underline{r} \qquad (7)$$

The n·n matrix $\underline{C}$ is the covariance-matrix of all reference values, the vector $\underline{f}$ contains the covariances between the reference points and the interpolation point. Formula (7) and formula (1) are identical, if we use as weight functions

$$\underline{w} = \underline{f}^T \cdot \underline{C}^{-1} \qquad (8)$$

The weight function of any reference point j is

$$w_j = \underline{f}^T \cdot \underline{C}^{-1} \cdot \underline{e}_j \qquad (9)$$

This formula shows, that the weight functions are comparable with the interpolation surface, if we use the standard values as reference values.

The weight functions of the least squares interpolation depend on the covariance function and the reference point distribution. The covariance function contains the filtering parameter f, which controls the fitting to the standard values. It also contains the parameter m, which is in fact a scale factor between the reference points and the covariance function. The distribution of the reference points can be regular or irregular. In figure 1 the influence of the covariance function parameters and of the distribution of reference points is shown.

The filtering parameter f describes the difference between the variance and the covariance for the distance $s = 0$. If the filtering parameter $f = 0$ is used, the weight functions will pass exactly through the standard values. The greater the magnitude of parameter f the greater the deviations from the standard values will be, and the interpolation surface will become smoother. The parameter m of the covariance function controls the appearance of the weight functions. We will get rapidly decreasing weight functions for small parameters m. Larger values of m give still decreasing, moderately swinging weight functions, until with large parameter values m we will get regularly curved weight functions in case of a regular reference point distribution, and greatly swinging weight functions for an irregular reference point distribution.

3. The characteristics of weight functions and their effects on interpolation surfaces

The different weight functions which appear in different interpolation methods can be compared and their characteristic features can be used for describing and classifying weight function types. Each type has its special effects on the interpolation surface whilst weight functions of the same type will cause similar interpolation results.

The following characteristics are used to define different types of weight functions and to describe their effects on the interpolation surface:

a) Independence of the reference values

A weight function can be dependent or independent from the reference values. If a weight function is independent from the reference values, the weight function (3) simplifies to

$$w_j = f_j(x_i, y_i ; u_1, u_2 \dots u_n ; v_1, v_2 \dots v_n) \qquad (10)$$

where u and v are the coordinates of the reference points.

A weight function which is independent of the reference values depends only on the distribution of the reference points, and the same weight functions will be obtained for any combination of reference values. Such weight functions describe the effects at the interpolation surface, if the reference values are varied. This is a very important interpretation of weight functions. It can be very helpful for imagining the interpolation surface in relation to the given reference points.

b) Standard values of the weight functions

The standard values have been defined in the preceeding chapter 2.2. With regard to the reference point j, we have

$w_j = 1$ for the reference point j

$w_j = 0$ for all other reference points.

If all weight functions pass through their standard values, the resulting interpolation surface will pass exactly through the reference points for any combination of reference values. If the weight functions don't pass through the standard values, the interpolation surface will in general not pass through the reference values (expect for some special sets of reference values). The expected differences between the reference values and the interpolation surfaces correspond to the deviations of the weight functions from their standard values. If the weight functions, wich satisfy the standard vàlues, are smoothed, the weight functions can only approximate the standard values. Those smoothed weight functions will result in smoothed interpolation surfaces with a good approximation to the reference values.

c) Normalization of the weight functions

Weight functions are called normalized, if the sum of all weight functions is 1 in each interpolation point. By using normalized weight functions, we will get a horizontal plane as interpolation surface, if all reference points have the same reference values. Otherwise deviations according to the sum of the weight functions would result.

d) Continuity and continuous derivatives of the weight functions

If this weight functions are not continuous or have not continuous derivatives, such characteristics will in general be found at the interpolation surfaces too.

e) Geometric features of weight function

In order to pass through the standard values, the weight functions must decrease from the value 1 in the respective reference point to the value 0 in the surrounding reference points. The essential geometric features of weight functions can be described by curvature characteristics and the resulting function values. The following classification will be used for the description of the geometric features of weight functions. In figure 2 examples are given for the various types of the weight functions and the associated interpolation results.

Regularly curved weights functions have quite regularly distributed and smoothly changing curvatures. The function values of the weight functions must not exceed much the value 1 and not become much smaller than the value 0. The descent from the value 1 to the values 0 will, for smoothness reasons cause small negative values beyond the surrounding reference points. With increasing distance from the reference point, the weight functions will approach 0. Regularly curved weight functions will cause interpolation surfaces with regular curvatures, as much as compatible with the reference values.

Linear weight functions decrease approximately linear from the value 1 in the respective reference point to the value 0 in the surrounding reference points. Beyond the closest surrounding reference points the weight functions will not differ very much from the value 0. The interpolation surfaces will have the greatest change of slope in the reference points, while the surface between the reference points is only affected by the closest surrounding reference points.

Rapidly decreasing weight functions decrease at a certain short distance from the reference point very fast from the value 1 to the value 0 which value is then kept. Such weight functions can describe terraced terrain forms with great slopes between them.

Swinging weight functions reach values much greater than 1 and smaller than 0. Such weight functions can cause unwanted interpolation surfaces, which may swing very much beyond the values of the reference points.

4. Weight functions as a linear combination of base functions

The interpolation with weight functions is justified and based on the insight that the characteristic features of an interpolation can be described and determined by weight functions. For different applications we can create adequate weight functions and we will get an interpolation surface according our specifications. The mathematical method for defining weight functions is free and can be derived from one of the known interpolation methods.

It is convenient not to design weight functions as such but rather to define them as linear combinations of certain more fundamental base functions. In this way all required types of weight functions can be created. The number of base functions is adapted to the number of reference points considered. In this way the weight functions can satisfy all standard values. Also smoothing of weight functions can be done to variable degrees.

4.1 Some base functions and the resulting weight functions

A base function is a function related to a certain reference point. The linear combination of the base functions, which are related to the different reference points, should keep the same number of surface parameters, so that the interpolation surface can not be replaced by a similar surface with less parameters. Adequate base functions are for instance

$$b_A = \sqrt{1+s^2/m^2} \qquad b_C = 1+s/m$$

$$b_B = \frac{1}{1+s^2/m^2} \qquad b_D = \frac{1}{1+s/m}$$

This base functions all contain the parameter s, which is the distance from the interpolation point to the respective reference point. The base function parameter m is used as a scaling parameter between the base function and the reference points.
After refering n base functions to the n given reference points, we get the weight functions w_j as

$$w_j = \underline{f}_j^{\,\prime} \cdot \underline{a}_j = b_{1j} a_{1j} + b_{2j} a_{2j} + \ldots + b_{nj} a_{nj} \qquad (11)$$

where the coefficients $\underline{a}_j$ are unknown.

By the condition, that the weight function w_j has to pass through the standard values, we get the equations

$1 = b_{1j} a_{1j} + \ldots + b_{nj} a_{nj}$ for the standard value of the reference point j

$0 = b_{1j} a_{1j} + \ldots + b_{nj} a_{nj}$ for all other standard values

The unknown coefficients $\underline{a}_j$ can be computed from this equation system $\underline{e}_j = \underline{F}_j \cdot \underline{a}_j$ and the weight function are obtained by

$$w_j = \underline{f}_j^T \cdot \underline{a}_j = \underline{f}_j^T \cdot \underline{F}_j^{-1} \cdot \underline{e}_j \qquad (12)$$

The linear combination allows the use of different base function parameters for each reference point. It allows also different base functions to be used. For each weight function different base functions can be referred to a reference point. But it is advantageous to relate always the same base function with the same parameter to a reference point, because then the interpolation surface can be computed directly as

$$z_i = \underline{f}^T \cdot \underline{F}^{-1} \cdot \underline{r} \qquad (13)$$

By using the 4 mentioned base functions with the same base function parameters, we get different types of weight functions depending on the base function parameters and the reference point distribution. There is one important difference between the base functions b_A, b_B and b_C, b_D, respectively, with regard to continuity of the derivatives. The base functions b_C, b_D do not have continuous derivatives and therefore the weight functions can break in each reference point.

For the base function $b_A = \sqrt{1+s^2/m^2}$ we get linear weight functions for small base function parameters, and regularly curved weight functions for medium size parameters. If we use large base function parameters, we get regularly curved weight functions for a regular distribution of reference points and greatly swinging weight functions for a irregular reference point distribution. The weight functions with the base function $b_B = 1/(1+s^2/m^2)$ are only different for small base function parameters. Instead of linear weight functions we get in this case rapidly decreasing weight functions. For the base function $b_C = 1+s/m$, the weight functions are always linear for all base function parameters and for all reference point distributions. For the base function $b_D = 1/(1+s/m)$ we also get linear weight functions. Only for small base functions parameters, the weight functions become decreasing.

4.2 Smoothing of weight functions

A weight function, which is a linear combination of base functions, can be smoothed by changing the coefficients of the linear combination. Reduced coefficients also give smaller derivatives of the weight functions. As the coefficients are computed under the condition, that the weight functions have to satisfy the standard values, smoothed weight functions can only approximate the standard values.

If we use an adjustment system, one part of the error equations can describe the derivatives of the weight function from the standard values as

$$\underline{v}_{Ij} = \underline{F}_j \cdot \underline{a}_j - \underline{e}_j \qquad (14)$$

and the other part of the error equations can describe the size of the unknown coefficients

$$\underline{v}_{IIj} = \underline{I} \cdot \underline{a}_j \qquad (15)$$

After referring the weitht matrix P_I to the deriatives at the standard values and the weight matrix P_{II} to the size of the coefficients, we get the least squares solution:

$$\underline{a}_j = (\underline{F}_j + (\underline{F}'\underline{P}_I)^{-1}\underline{P}_{II})^{-1} \cdot \underline{e}_j \qquad (16)$$

A formal consideration of (16) shows, that smoothed weight functions are obtained, if to the matrix $\underline{F}$ of (12) the Matrix $\Delta\underline{F} = (\underline{F}_j'\underline{P}_I)^{-1}\underline{P}_{II}$ is added. Under certain assumptions, the matrix $\Delta\underline{F}$ can be reduced to a diagonal matrix. It can be shown, that changing the diagonal elements of the matrix $\underline{F}$ in (12) achieves an effective smoothing of the weight functions. Thorough investigation established the following rules concerning the modification of the diagonal elements of matrix $\underline{F}$:

a) For the base functions $b_A \sqrt{1 + s^2/m^2}$ and $b_C = 1 + s/m$, the magnitudes of the diagonal elements have to be reduced, for the base functions $b_B = 1/(1 + s^2/m^2)$ and $b_D = 1/(1 + s/m)$, they have to be increased, in order to get smoothed weight functions.

b) The degree of smoothing is dependent on the magnitudes of modification.

c) The weight functions will pass exactly through the standard values, as long as the initial diagonal elements of $\underline{F}$ are not changed.

5. The description of terrain surfaces by interpolation with weight functions

The interpolation with weight functions can be used for highly qualified description of terrain surfaces by sophisticated DTM. The difficulties of the description of terrain surfaces are: the often irregular distribution of the reference points, the consideration of additional terrain information and a qualified smoothing of the surface. By specifying the features of the expected interpolation surface, it is possible to establish the required weight functions and to obtain the wanted features.

In the following it will be shown, how interpolation with weight functions can be used to obtain a high accuracy description of terrain surfaces. Three representative examples of such interpolation results are presented in fig. 3.

5.1 Irregular distribution of the reference points

For the description of terrain surfaces, regularly curved weight functions have to be used. Linear weight functions will cause lower accuracy, as the 2nd order neighbouring reference points have to be considered for a sufficient description of the surface. However, it is advantageous, if the weight functions tend to linear functions rather than to greatly swinging functions.

Regularly curved weight functions with a tendency to approximate linear functions will be obtained, if we use the base function $b_A = \sqrt{1+s^2/m^2}$ and let the parameter m vary for each base function according to the minimum distance of the reference point to the closest next point. The weight functions are to be normalized, as described above (chapter 3).

5.2 Consideration of additional information

Additional information about terrain surfaces can be available in many ways. Very often it is given by special lines or points. Such lines are for instance lines of extreme curvature, lines of steepest descent, or break lines. As special points are used the local highs and lows of the terrain. The following remarks will explain the information which such lines and points represent. Break lines, which also represent important information about the terrain, will not be further discussed is this paper.

Curvature lines: Across such lines the terrain slopes are changing. These lines can be used for describing the boundaries between tilted planes without being break lines proper.

Lines of steepest descent: Perpendicular to such lines the terrain surface has horizontal tangents.

Highs and lows: The highs and lows locate a local maximum or minimum of the terrain surface. They imply that the respective tangential planes must be horizontal.

Curvature lines are considered, if the density of the reference points along the lines is increased with respect to the density of the other reference points. The distances to the surrounding points will become smaller and the weight functions will closer approximate linear functions. For taking lines of steepest descent into consideration we have to use supplementary functions. In addition to the base function, we also relate a supplementary function to each point of those lines. For each supplementary function, we get a new free coefficient a. The coefficients a are computed in such a way, that the interpolation surface passes through the reference points and that the 1st derivativesof the surface perpendicular to the descent lines disappear. The function

$$b_s = \frac{tx}{1 + s^2/m^2}$$

can be used as supplementary function, referring to local coordinate systems wich relate to the points of the descent lines.

Considering highs and lows, we have to relate 2 supplementary functions to each point. Then the condition can be satisfied, that the 1st derivatives in two orthogonal directions are 0.

5.3 The smoothing of terrain surfaces

The smoothing of terrain surfaces can have two different purposes. One purpose can be the filtering of the reference values. Their errors (scanning errors, local terrain irregularities) are not to be superimposed on the interpolation surface. The second purpose of smoothing is generalisation. The problem of generalisation is that certain lines and groups of points should be maintained. The interpolation surface should still pass exactly through those points and lines, which describe the most characteristic features of the surface. The original curvatures should be kept too. Only between those points and lines smoothed surfaces are expected.

For smoothing, we increase the base function parameters. We can either introduce minimum parameter values or multiply the initial parameters by a factor. In order to get smoothed weight functions and to avoid greatly swinging functions, certain corrections Δf have to be subtracted from or added to the diagonal elements of the matrix $\underline{F}$, see chapter 4.2. The magnitudes of Δf depend on the increased parameters m_i and the distance s_m to the next reference point. A suited function was found to be

$\Delta f = 0.05\ (1 - s_m{}^2/m_i{}^2)$

Filtering of reference values is obtained by using minimum parameter values. The interpolation surface will still pass through all reference points, which have no neighbouring points within a distance smaller than the minimum parameter. If there is a nearer neighbouring point, the respective element of the matrix $\underline{F}$ will be changed, and the surface will not pass exactly through this reference point. The actual deviations depend on the reference values.

The influence of systematic deviations between different recording units can be considered, if the parameters of the base functions are only computed within one unit. If there is a closer point from an other recording unit, smoothing will take place.

Scanning errors can be eliminated if we use different scales in the direction of and orthogonal to the profiles. The distance between two profiles is reduced to half, in order to have a regular grid for all profiles of the same scanning direction. The grids of both scanning directions are superimposed with a shift of half of the grid width. Then a smoothed interpolation surface can be computed accordingly like for two different recording units. The result is transferred back to the original system.

For generalisation we use minimum parameters of the base functions or multiply the parameters by a factor. The principle is to increase the base function parameters in order to smooth the surface. For the special points and lines, as mentioned, the original parameters will be kept and the interpolation surface will still pass exactly through those reference points. Also the original curvatures will not be changed very much.

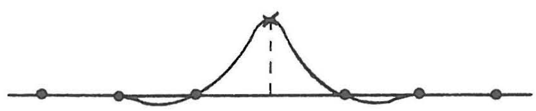
5.4. Examples

Three contourline maps are presented in figure 3. They were computed with the modified SCOP program where the least squares interpolation was replaced by the described interpolation method with weight functions. For all three examples, the same reference points have been used, and the contour interval is 5 m. In example 1 all reference points are irregularly distributed points. The example 2 shows the terrain surface, if a classification is made between irregularly distributed points, points of curvature lines and descent lines, and highs and lows. Now the contourlines cross the lines of steepest descent rectangularly, and the highs and lows are enforced to be the highest or the lowest points of the surrounding surface. While in example 1 and 2 the surfaces pass exactly through the reference points, example 3 shows the effect of smoothing. The surface still passes exactly through the points of the curvature lines and the descent lines and also through the highs and the lows. In between, however, the surface is smoothed.

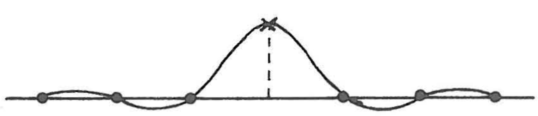
Fig. 1 Examples of weight functions of the least squares interpolation

Dependency of base function parameters

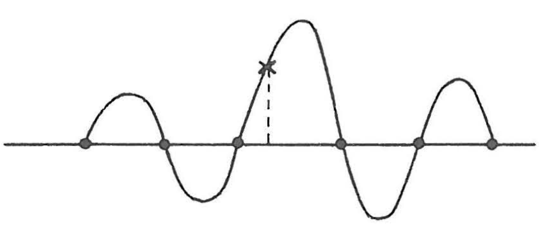
a.) rapidly decreasing weight function (parameter value small)



b.) regularly curved weight function (medium magnitude parameter)

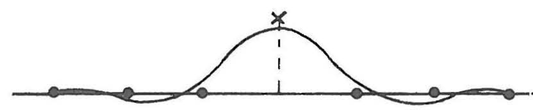


c.) greatly swinging weight function by irregular reference point distribution

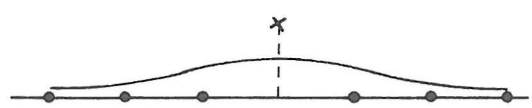


Dependency on filter parameters

a.) small filter parameter



b.) large filter parameter



c.) medium magnitude filter parameter, and irregular reference point distribution

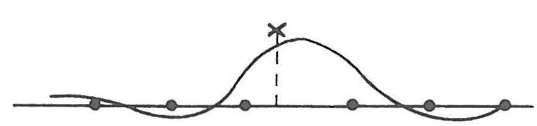


Fig. 2 Types of weight functions and related characteristic interpolation results

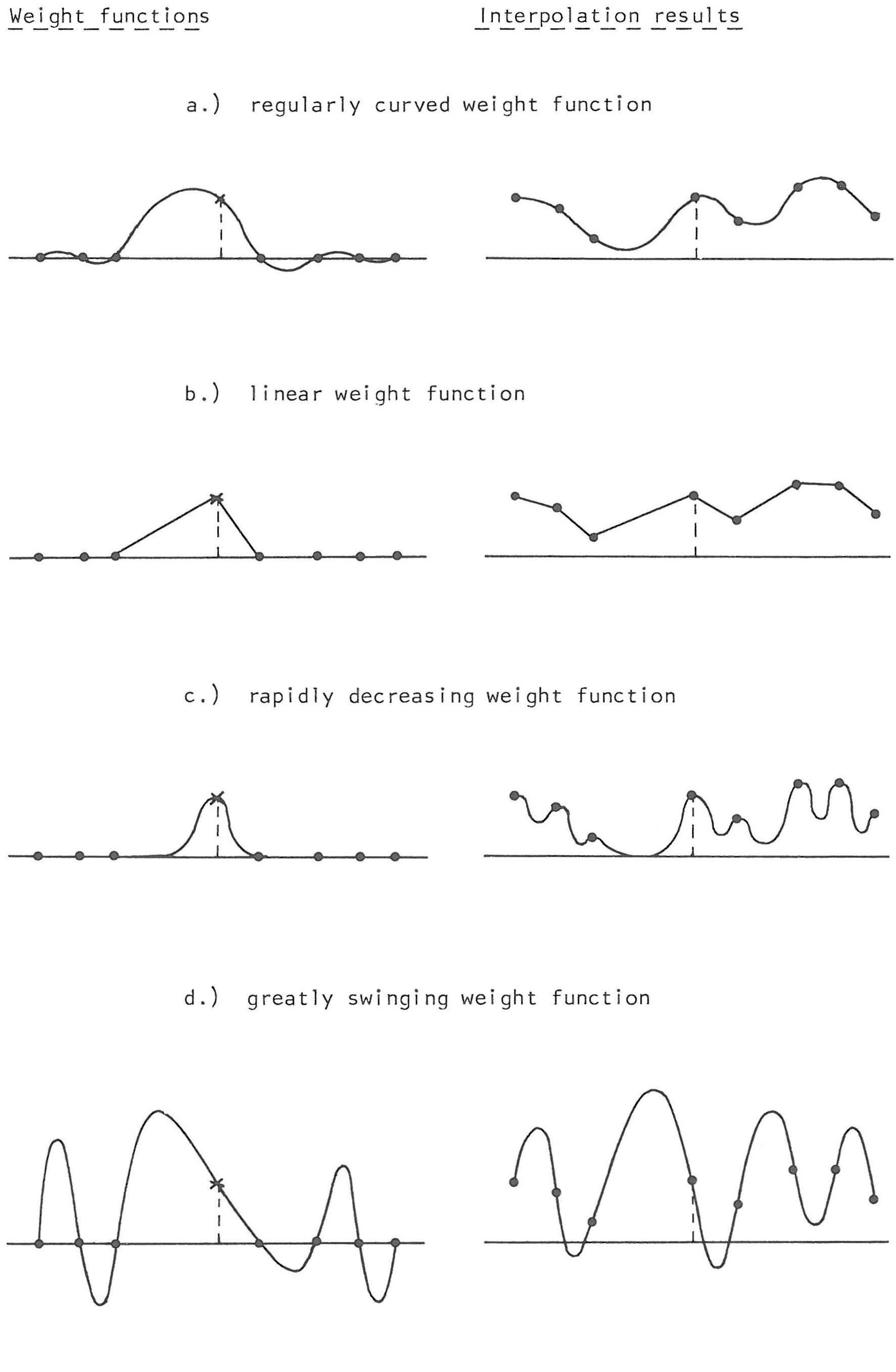
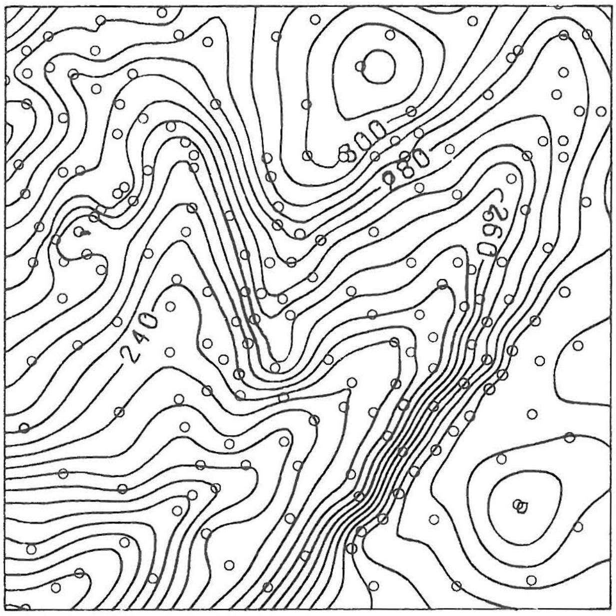


Fig. 3 Examples of interpolation with weight functions

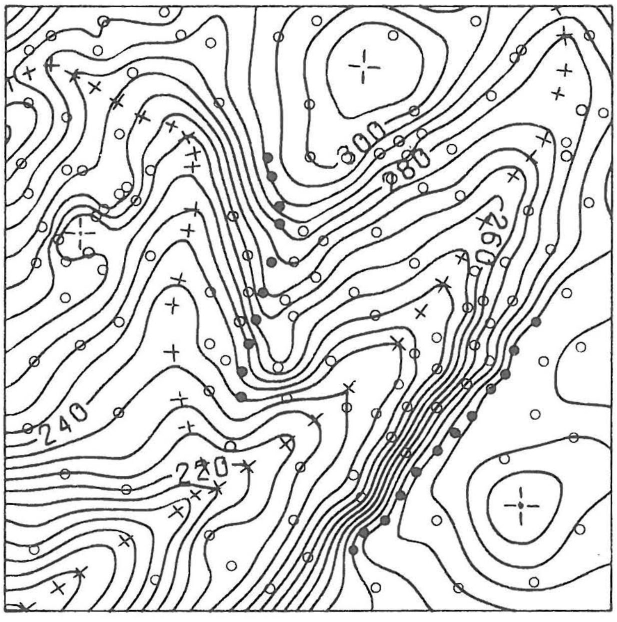
Classes of points used:

- o irregularly distributed points
- ● points on "curvature lines"
- + points on lines of steepest descent
- -¦- highs and lows

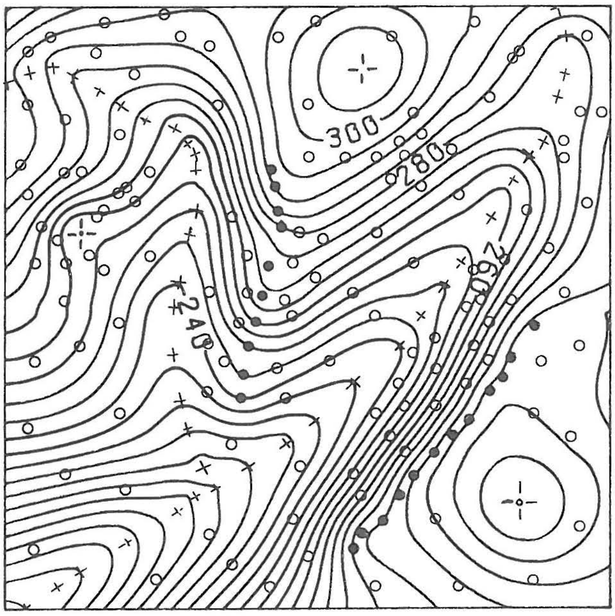
Example 1. irregular point distribution, no classification



Example 2. with classification of points, according to the characteristic terrain features



Example 3. With classification of points and smoothing



References:

(1) Assmus E.: Extension of Stuttgart Contour Program to treating terrain breaklines, presented paper, Comm. 3, XIII ISP Congress Helsinki, 1976

(2) Bosman E. R., Eckart D., Kubik K.: Delft - a programme system for surface approximation, Bildmessung und Luftbildwesen 40, 1, 1972

(3) Ebner H., Reiss P.: Height interpolation by the method of finite elements, presented paper, ISP Commission 3 - Symposium, Moscow, 1978

(4) Hardy R. L.: Analytical topographic surfaces by spatial intersection, Photogrammetric Engineering 38, 5, 1972

(5) Kraus K.: Ein allgemeines digitales Geländemodell, Numerische Photogrammetrie, Sammlung Wichmann, Band 5, 1973

(6) Shut G. H.: Review of interpolation methods for digital terrain models, Invited paper, Commisssion 3, XIII ISP Congress Helsinki, 1976

(7) Stanger W.: The Stuttgart Contour program SCOP - further development and review of its application, presented paper, Comm. 3, XIII ISP Congress, Helsinki 1976