(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 September 2011 (01.09.2011)

PCT

(10) International Publication Number
WO 2011/106527 A2

(51) International Patent Classification: *G06F 17/18* (2006.01)

(21) International Application Number: PCT/US2011/026078

(22) International Filing Date: 24 February 2011 (24.02.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 61/338,815 24 February 2010 (24.02.2010) US

(71) Applicant *(for all designated States except US)*: AMPERSAND INTERNATIONAL, INC. [US/US]; 23775 Commerce Park Road, Beachwood, OH 44122 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KALAMBET, Yuri [RU/RU]; P.O. Box 27, Moscow, 123182 (RU). MALTSEV, Sergey [RU/RU]; P.O. Box 27, Moscow, 123182 (RU).

(74) Agent: STEIN, David, J.; Cooper Legal Group, LLC, 6505 Rockside Road, Suite 330, Independence, OH 44131 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2011/106527 A2

(54) Title: METHOD FOR NOISE FILTERING BASED ON CONFIDENCE INTERVAL EVALUATION

(57) Abstract: One or more methods and techniques are disclosed for noise filtering based on confidence interval evaluation. A desired (e.g., minimal) confidence interval is used as a criterion for the selection of proper parameters for an approximating function comprising a degree of a polynomial, a number of points used in the approximation, and a positioning of a window center with respect to an approximated point for the polynomial approximation. The approximating function may be measured with error by the polynomial or other functions which allow for the estimation of the confidence interval. Confidence interval evaluation based noise filtering enables accounting for a priori information regarding the noise level of the signal.

METHOD FOR NOISE FILTERING BASED ON CONFIDENCE INTERVAL EVALUATION

RELATED APPLICATION

This application claims the benefit and priority at least in part of U.S. Provisional Patent Application No. 61/338,815, filed on February 24, 2010, entitled "Method of Noise Filtering Based On Confidence Interval Evaluation", which is herein incorporated by reference in its entirety.

BACKGROUND

[0001] Measured signals are comprised of a signal portion and a random error portion, which may be caused by a variety of factors, such as from electronic components, a variation of surrounding ambient conditions, and/or radio interference. It may be beneficial to minimize the error portion of the signal as much as possible in order to achieve an accurate estimate of the measured signal. Such noise reduction techniques may be based on polynomial approximation performed by linear smoothers (*e.g.,* Savitzky-Golay (SG), Fourier transform based convolution, convolution with Gaussian, and/or moving average) or a non-linear noise manner (*e.g.,* median filtering). However, such methods often produce results that may be unsatisfactory due to a low noise suppression and/or an undesirable influence on object boundaries or shapes. Results based on traditional noise filtering in chromatography or capillary electrophoresis may produce shape distortion and/or disturbances, which may be especially pronounced near sharp edges, baseline steps, and/or triangular peaks.

SUMMARY

[0002] This Summary is provided to introduce a selection of concepts in a simplified form which are further described below in the Detailed Description. This Summary is not intended to identify key factors or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

[0003] One or more methods, techniques, and/or systems are disclosed for a family of noise filtering methods based on a variety of approximating functions.

[0004] According to one aspect, the confidence interval evaluation noise filtering is implemented by way of a polynomial filter, wherein a signal comprising an array or vector of original signal values corresponding to points within the signal may be evaluated using the confidence interval evaluation noise filtering technique. In one embodiment, the polynomial may have a fixed-degree, while other embodiments may comprise varying degree polynomials. In accordance with techniques presented herein, approximation polynomial is constructed using least-square method and approximated values of a signal may be evaluated as a set of points within an approximation window of the signal, along with a corresponding confidence interval. In contrast to the Savitzky-Golay (SG) technique, where an approximated value in the center of a window is used, the technique(s) presented herein involve re-evaluating the set of approximated values when the approximation window shifts by a point, along with new confidence intervals for the corresponding points. An approximated value for the signal may be compared with the corresponding value from the output array of values, and depending on the confidence interval for that corresponding point, the approximated value may replace the stored estimate of signal value or alternatively, the stored approximation may be kept.

[0005] In one embodiment, the approximated value of the signal replaces the original signal value within the array or vector if the confidence interval of the new signal value is lower than the current confidence interval.

[0006] By application of the techniques presented herein, confidence interval evaluation based noise filtering may be based at least in part on a least squares approximation and measurement signals, comprising objects that may be recoverable from noise by calculations utilizing confidence intervals. Confidence interval evaluation noise filtering may be repeated iteratively, by repeatedly selecting different sets of approximating functions to improve overall accuracy. Therefore, confidence filtering provides desired (*e.g.,* minimal) confidence intervals for respective approximation points.

[0007] In one embodiment, polynomials may be accepted or rejected based on the standard deviation σ of the noise portion of the signal measurement. For example, a polynomial exhibiting a low probability of fitting approximated data properly and/or accurately may be rejected due to poor modeling, and other polynomials may be selected for improved confidence interval estimates. Such confidence interval evaluation based noise filtering may be applied to a variety of measurement scenarios, including but not limited to the fields of analytical chemistry, chromatography, spectrometry, and/or image processing.

[0008] To the accomplishment of the foregoing and related ends, the following description and annexed drawings set forth certain illustrative aspects and implementations. These are indicative of but a few of the various ways in which one or more aspects may be employed. Other aspects, advantages, and novel features of the disclosure will become apparent from the following detailed description when considered in conjunction with the annexed drawings.

DESCRIPTION OF THE DRAWINGS

[0009] The application is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

[0010] Figure 1 is an illustration of a flow diagram illustrating an example embodiment where one or more techniques described herein may be implemented.

[0011] Figure 2 is an illustration of a flow diagram illustrating an example embodiment where one or more techniques described herein may be implemented.

[0012] Figure 3 is an illustration of a flow diagram illustrating an example embodiment where one or more techniques described herein may be implemented.

[0013] Figure 4 is an illustration of an example comparison between SG and confidence interval filtering.

[0014] Figure 5 is an illustration of an example approximation for the G_1 value of Figure 4.

[0015] Figure 6 is an illustration of an example approximation for the G_2 value of Figure 4.

[0016] Figure 7 is an illustration of an example comparison between SG and confidence interval filtering as applied to electrophoresis data.

[0017] Figure 8 is an illustration of an example chromatographic analysis.

[0018] Figure 9 is an illustration of an example comparison between SG and confidence interval filtering.

[0019] Figure 10 is an illustration of an example comparison between SG and confidence interval filtering.

[0020] Figure 11 is an illustration of an example comparison between SG and confidence interval filtering.

[0021] Figure 12 is an illustration of an example comparison between SG and confidence interval filtering.

[0022] Figure 13 is an illustration of an example series of peaks.

[0023] Figure 14 is an illustration of an example comparison between a filtered signal and a pure signal.

[0024] Figure 15 is an illustration of an example comparison between SG and confidence interval filtering as applied to electromyography (EMG).

[0025] Figure 16 is an illustration of an example of pump pulsations.

[0026] Figure 17 is an illustration of an example comparison between SG and confidence interval filtering as applied to capillary electrophoresis (CE).

[0027] Figure 18 is an illustration of an exemplary computing environment wherein one or more of the provisions set forth herein may be implemented.

[0028] Figure 19 is an illustration of an exemplary computer-readable medium comprising processor-executable instructions configured to embody one or more of the provisions set forth herein.

DETAILED DESCRIPTION

[0029] The claimed subject matter is now described with reference to the drawings, wherein like reference numerals are generally used to refer to like elements throughout. In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the claimed subject matter. It may be evident, however, that the claimed subject matter may be practiced without these specific details. In other instances, structures and devices are shown in block diagram form in order to facilitate describing the claimed subject matter.

[0030] One or more methods, techniques, and/or systems are disclosed for calculating an approximated value for a signal based on adaptive approximation using confidence intervals. Approximating functions are selected using criterion based on desired (*e.g.*, minimal) confidence intervals of the approximation. For example, one such function capable of confidence interval calculation is a polynomial approximation using the least squares method, as will be discussed in greater detail herein. As an illustrative example, Figure 7 (discussed in greater detail below) demonstrate exemplary advantages of confidence interval filters by applying a third-degree adaptive polynomial filter to a set of chromatographic data and comparing the results with the popular Savitzky-Golay (SG) filtering method using a polynomial of the third degree. Therefore, the following comparisons, illustrations, and figures are based on confidence interval evaluation noise filtering compared with the SG filtering method, which is frequently used and currently considered one of the most reliable methods available.

[0031] Confidence interval evaluation based noise filtering has several advantages over other methods of noise filtering, and among others, an automatic adaptive reduction of the approximation window for accurate approximation of an object (*e.g.,* a peak or a baseline region). Therefore, the shape of the peak within a signal does not suffer when a confidence interval filter is applied by comparison to other noise filtering methods. Confidence interval filtering effectively suppresses baseline noise, significantly improves the detection and quantification limits, and even allows for the suppression of non-white noise (*e.g.*, pump pulsations discussed in greater detail below, chemical noise, *etc.*) Further, confidence interval evaluation provides the advantages of non-central approximations of the signal in close proximity to object boundaries (*e.g.,* a baseline step or a peak upslope). Improper approximation for known noise levels may also be avoided by using confidence interval evaluation due to the fact that noise filtering achieved through confidence interval evaluation may be based on the premise that the quality of a polynomial approximation may not be uniform through an analyzed signal over the entire range of the signal. According to one aspect, statistical methods and regression analysis enable a definition of a numerical evaluation of the quality of the approximation. Thus, a natural evaluation of the quality of an approximation is a confidence interval for an approximated value.

[0032] To this end, noise filtering based at least in part on confidence interval evaluation may pertain to a variety of measurement methods and systems, including but not limited to measurements related to analytical chemistry, chromatography, spectrometry, combined techniques, and/or image processing. A number of other applications or examples may be readily imagined (*e.g.*, accurate handling of an object's boundaries in image processing). Further, confidence interval filtering may be applied to isocratic chromatography, where the peaks at the beginning are narrow, while the peaks tend to broaden toward the end of the chromatogram. Conventional techniques may make it difficult to filter such a signal correctly, since a small gap on a conventional filter produces low noise suppression, while a bigger gap produces distortions of narrow peaks. Additionally, adaptive confidence filters may resolve objects of various widths or different frequencies in a Fourier domain which may be present simultaneously in a signal.

Variable Definitions

[0033] For the purposes of discussion, the following example is based on single dimensional data (*e.g.*, a chromatographic signal) and homoscedastic noise. However, the techniques described herein may also be applied to two or multi-dimensional data filtering (*e.g.*, image processing). Further, this single dimensional approach may be repeated using different single-dimensional subsets of multi-dimensional data arrays as desired.

[0034] The input array comprises an array of raw data, while the elements of the output array comprise an approximated data point and an estimate of the confidence interval corresponding to the point. Accordingly, definitions for variables and equations are outlined below:

$$\mathbf{Y} = \{ y_1, y_2 ..., y_n \}$$

represents a vector of detected response values from a signal.

$$\hat{\mathbf{Y}} = \{ \hat{y}_1, \hat{y}_2 ..., \hat{y}_n \}$$

represents a vector of estimated and/or approximated response values calculated based on the detected response values and least squares regression. The least squares approximation may be calculated using the formula below:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$ and

the general definition of matrix **X** is

$$\mathbf{X} = \begin{Bmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) & \ldots & f_p(x_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1(x_n) & f_2(x_n) & f_3(x_n) & \ldots & f_p(x_n) \end{Bmatrix}$$

Where $f_k(x_i)$ are the values of approximating functions. In the case of polynomial approximation

$$f_k(x_i) = x_i^{k-1}$$

And matrix X becomes

$$\mathbf{X} = \begin{Bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{p-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{p-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 & \ldots & x_n^{p-1} \end{Bmatrix}$$

where $\mathbf{X}$ represents a matrix of values (as shown above) which runs along an independent axis (*e.g.*, a time axis or a position axis).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{X}' \cdot \mathbf{Y}$$

where $\hat{\boldsymbol{\beta}}$ represents polynomial coefficients for regression.

$$\mathbf{x}_*' = \{ f_1(x_*), f_2(x_*), \ldots, f_p(x_*) \}$$

and for the case of polynomial approximation

$$\mathbf{x}_*' = \{1, x_*, \ldots, x_*^{p-1}\}$$

where x_* represents a position at which smoothed and/or approximated values may be estimated.

$$C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$$

Here, C_y represents the calculated confidence interval with respect to a corresponding point from the vector of detected response values, or in other words, a corresponding point for the measured signal. Further, this confidence interval is based on homoscedastic noise rather than heteroscedastic noise. In some heteroscedastic noise cases, the raw data may be scaled to avoid complications. For example, one may take the square root of a signal before smoothing, and use square of the value in output array result as an estimate of the initial signal.

Unlike SG, the confidence interval may be calculated for any x_* within an approximated range for a given polynomial approximation (while SG relies on a central point). Further, original data is not required to be uniformly distributed along the position axis in order to calculate a confidence interval. Therefore, the positioning of x_* for the approximated value is not required to coincide with a point where initial data was defined. Confidence intervals techniques may smooth irregular data accordingly and enable re-sampling of regular and irregular data.

$$t_m^{\delta}$$

Accordingly, the Student's coefficient for confidence level (1 - δ), where m represents the number of degrees of freedom and δ represents a P-value.

In the formula of C_y:

α = (1 - confidence level)

where confidence level may be defined within the range (0, 1). As an example, if a confidence probability value of 0.95 was selected, then α would be calculated to be (1 – 0.95) = 0.05

n = the number of data points used for the polynomial approximation (*e.g.*, the gap of the filter)

p = number of parameters of the polynomial, power plus one

m = n - p = the number of degrees of freedom

σ^2 = noise variance

σ = noise standard deviation

RSS = residual sum of squares

S^2 = an estimate of the noise variance,

$$S^2 = \frac{RSS}{n-p}$$

$$RSS = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$S^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-p}$$

$$u_* = \mathbf{x}_*' (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{x}_*$$

Confidence Interval Formulae

[0035] As an illustrative example, if there is a case of a large data array with N data points, an estimate S^2 of the population noise variance σ^2 may be obtained from a local subset of n points. However, this estimate may not be particularly accurate in certain circumstances, especially in the case of a small number of degrees of freedom m = (n-p).

[0036] Therefore, the formula for the confidence interval estimate can be rewritten from:

$$C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$$

to

$$C_Y = t_{n-p}^{\frac{1}{2}\alpha} \cdot \sigma \cdot \sqrt{u_*}$$

by replacing the S with σ.

The population noise variance σ^2 may be represented as follows:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \tilde{y}_i \right)$$

where $\tilde{y}_i$ is a "true" signal value at position *i.*

Additionally, different confidence interval C_y calculation rules may be implemented by several different formulae based on the value of S^2.

For example,

$$C_Y = \begin{Bmatrix} S^2 < \sigma^2, Formula\,(a) \\ S^2 \geq \sigma^2, Formula\,(b) \end{Bmatrix}$$

Formula (a) $C_Y = t_{n-p}^{½\alpha} \cdot \sigma \cdot \sqrt{u_*}$

Formula (b) $C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$

Therefore, if S^2 is less than σ^2, Formula (a) may be applied, while if S^2 is greater than or equal to σ^2, Formula (b) may be used.

[0037] The noise variance may also provide the benefit of evaluating the quality of an approximation model. When an approximation model is correct, a larger number of points *n* in an approximation window may result in an estimate of variance S^2 which is closer to the true value of σ^2. In an example based on white homoscedastic noise values $(n-p)\ S^2$ should have a χ^2 distribution, therefore the engine of hypothesis testing for evaluation of the model quality may be used. Accordingly, the hypothesis that an approximation fits a set of data and is statistically significant needs to be verified (*e.g.*, typically decided using a $1-\alpha = 0.95$ confidence cut-off level, but other values may be used depending on data, and this cut-off level may be used as one of parameters of the filtering algorithm).

[0038] However, if the hypothesis fails, the polynomial may be treated as having a systematic error (which may include but is not limited to: an excessive number of points used to construct the polynomial, an insufficient power for the polynomial (*e.g.*, the degree of the polynomial is too small), or a polynomial not fit for approximating the data). Upon failure of the hypothesis:

$$RSS_i \geq \sigma^2 \cdot \chi^2(\alpha, n-p)$$

$\chi^2(\alpha, m)$ is a χ^2 distribution function;

α represents the P-value; and

m represents the number of degrees of freedom.

Evaluating Noise Level

[0039] An estimate of σ^2 may be made using a data array, measured with a constant data rate, although alternative algorithms may be used in other cases. As an example, a robust method of noise estimation may be selected by defining a width of the noise window n and a degree of the polynomial $p-1$, Noise Definition Polynomial (NDP) to approximate the data.

[0040] In one embodiment, most of the signal peculiarities (which are treated as the noise of the data) are effectively suppressed if the array is filtered by SG based on the width of the noise window. Likewise, most of the model's noise-free data would be properly described by the polynomial of the associated window and/or power. Accordingly, the data array may be approximated numerous times, starting from the beginning of the array, and repeated when the approximation window is shifted to higher index points n/3. Likewise, at the respective approximation positions, the residual sum of squares *RSS* may be calculated and stored to the vector **E**. In one embodiment, vector **E** may be used for estimating average *RSS* values by way of iterative look-through passes.

[0041] For example, during the first pass, an average of values in **E** may be calculated, and the first estimate for σ_0^2 may be calculated as follows:

$$\sigma_0^2 = \frac{1}{M \cdot (n-p)} \sum_{i=1}^{M} E_i$$

M = the number of elements in the array **E.**

[0042] On subsequent passes, calculations for new improved estimates may be accepted upon meeting the condition:

$$E_i \leq \sigma_{j-1}^2 \cdot \chi^2(0.05, n-p)$$

and new σ_j^2 values may be calculated.

[0043] Continuing with this example, the outlier rejection process may be repeated until σ^2 stops changing, but generally no more than 5 times, as the procedure effectively rejects all outliers originating from regions with a poor approximation of the data (*e.g.*, baseline steps, jumps, and/or sharp peaks caused by sample injections). Therefore, an estimate of σ^2 may be obtained in the last step of iterations, and accepted as the noise σ^2 for the entire array.

Confidence Filtering

[0044] It will be appreciated that the acts described below are merely examples and the method(s) described herein may be practiced with more or fewer acts and/or in an order different from that described and/or illustrated in the figures.

[0045] Different implementations of the filtering and re-sampling functions are possible depending on the nature of the analyzed signal. Techniques may be implemented where variations of the gap (n), power of the polynomial (p), and the non-central approximation are combined in different ways. In one embodiment, the gap (n) and power of the polynomial (p) are fixed and optimal approximations may be chosen from all possible non-central approximations for a given position x_*. In another embodiment, the power of the polynomial (p) is fixed while various sets of gaps (n) are used. An optimal approximation is selected from all possible non-central approximations and from all gaps in the set for a given position x_*. In yet another embodiment, the gap (n) is fixed while power of the polynomial (p) may be varied within a range of [1, p]. Optimal approximations may be chosen from possible non-central approximations and from all the powers for a given position x_*.

[0046] Figure 4 illustrates an example where gap (n) and power (p) are fixed, and the signal comprises a baseline, a section of a peak, and a steep slope. G_1 and G_2 illustrate positions of approximation intervals for evaluation of point x_*. Assuming that data smoothing may be desired, the power of the polynomial is p = 3. Figure 4 illustrates how G_1 and G_2 may be used for approximating a response value at point x_*. ($G_1 = G_2 = 61$)

[0047] Figures 5 and 6 illustrate approximations at G_1 and G_2 for Figure 4, where confidence intervals at a 0.95 confidence probability level are used. Approximations at G_1 are based on conventional central SG filtering, while approximations at G_2 are non-central. This gives a better approximation closer to the pure signal while maintaining a confidence interval less than the one associated with the central approximation. In Figure 6, non-central approximation G_2 produces efficient results for noise suppression analogous to a pure baseline. Central approximation G_1 produces a significant amount of distortion. Ultimately, confidence intervals for both cases provide the criteria to decide which approximation may be better.

[0048] As stated above, the confidence interval calculations may be applied to various noise reduction tasks. Therefore, any subsets of points comprising the raw data array may be approximated by a polynomial. Multiple estimates of the values and confidence intervals at the target position may be made using different data subsets (*e.g.,* different sizes, widths, and/or positions for the approximation window). Additionally, multiple estimates may be calculated using varying degrees for the polynomials.

[0049] In an example embodiment illustrated by Figure 1, confidence filtering may be implemented in a method 100 based on the following description. At 102 σ^2 is evaluated for the noise distribution using the data array (alternatively, the evaluation may be optional if noise distribution parameters are known in advance or are being ignored). At 104 a list of polynomial construction parameters (*e.g.*, elements) to be applied is defined where the respective elements within the list comprise a polynomial degree and a subset of input data used to build a polynomial or data subset construction rule (*e.g.*, window width, position, size). At 106 the data is approximated based at least in part on data within the approximation window by the polynomial according to parameters of the respective list elements. At 108 the RSS is evaluated based on the following conditions 110:

Noise Not Ignored

[0050] If noise is not being ignored and the RSS is too big and also satisfies the following condition:

$$RSS_i \geq \sigma^2 \cdot \chi^2(0.05, n-p)$$

then the input data arrays in the vicinity of the output positions which did not get estimates are analyzed.

Noise Ignored

[0051] If noise is being ignored, the RSS is sufficiently small, or the RSS does not satisfy the aforementioned condition, then a new confidence interval is calculated 112 based on one or more of the following formulae:

(a) $C_Y = t_{n-p}^{½\alpha} \cdot \sigma \cdot \sqrt{u_*}$; or

(b) $C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$; (noise is ignored)

(c) $C_Y = \begin{Bmatrix} S^2 < \sigma^2, Formula\,(a) \\ S^2 \geq \sigma^2, Formula\,(b) \end{Bmatrix}$ (noise is not ignored)

[0052] For respective points within the window, the new confidence intervals are compared 114 with the values at the corresponding output array. The new confidence interval and the corresponding approximation point may replace the old value stored in the output array if the condition that the new confidence interval is smaller than the old value is satisfied. Approximations may be made until each element of the list has been run 116. At 118 positions which were not evaluated are treated as outliers and data around these positions is re-evaluated. For example, if the noise is not homoscedastic or is not an outlier, then a portion of the input data array may not be analyzed. Additionally output data may be filled with estimates based on revised noise parameters or other principles.

[0053] In another embodiment, approximations may be rejected based on contradictions with a priori information related to the raw data array (*e.g.*, noise level). Therefore, the approximation comprising the window size, position, and polynomial degree with the desired (*e.g.*, minimal) confidence interval may be accepted as a correct approximation for a target position.

[0054] Additionally, confidence intervals may be extended to data of higher dimensions (*e.g.*, 2-D, by approximation of smoothed data in a different dimension). The heteroscedastic noise model may be a benefit for the second dimension since regression may gain additional accuracy by accounting for different confidence intervals of data points.

Example Filtering Using Fixed Window And Fixed Degree Polynomial

[0055] The following example demonstrates chromatographic data measured with a constant data rate, where x is an integer (*e.g.*, the index of data array values). Non-central approximations may be especially effective in chromatographic data analysis where detector wavelengths or sensitivities may be switched between measurements. In this example, re-sampling is not considered, and therefore the smoothed values of measured data points may be calculated, while the response values for non-integer x values may not be calculated.

[0056] Figure 2 illustrates a method 200 where noise filtering may be achieved by evaluating points and confidence intervals 202 for respective points within a selected window based on a polynomial approximation. For the respective points within the window, the new confidence interval may be compared 204 with the corresponding output array value. If the new confidence interval is smaller than the value stored in the output array, or if a point is not evaluated, then the point stored in the output array may be replaced, along with its corresponding confidence interval. Therefore, the evaluation window may be shifted 206, and new points and confidence intervals may be evaluated by repeating from 202. Accordingly, each point may be approximated n times and an estimate with the best confidence interval may be selected as a filtered value. This method is generally more effective for a reasonable high number degrees of freedom (*e.g.*, $m = n - p >$ 4). Artifacts may be caused by a poor estimate of the local noise level, which may lead to an incorrect estimate of the confidence interval.

[0057] Figure 3 illustrates an exemplary method 300, which begins at 302 and involves executing the acts of 304 on the processor instructions configured to perform the techniques presented herein. It will be appreciated the acts of exemplary method 300 may be based at least in part on the formulae presented herein (*e.g.*, ***Variable Definitions, Confidence Interval Formulae***). Exemplary method 300 demonstrates approximating a set of one or more approximated values for a signal based at least in part on receiving a signal comprising noise and an input data array of three or more signal values which correspond to a set of three or more points, the points comprising an independent X coordinate and a measured Y coordinate, as shown at 302. Further, as part of 302, the signal may be scaled with a scaling function and/or an inverse scaling function. Additionally, receiving the signal 302 may comprise sorting the input data array in an ascending order based at least in part on at least one of the X coordinates. At 304 an output data array is approximated, comprising a number of elements based at least in part on a selected number of targets to approximate, the respective elements comprising an approximation and a confidence interval associated with the respective approximation, the approximation comprising an X coordinate and a Y coordinate, the number of targets defined by a nature of a task. The subset of the input data array of 304 may be selected based at least in part on a defined number (approximation window, *n*) of adjacent input points. Further 304 may comprise a confidence level based at least in part on a user-defined parameter of the algorithm. The instructions comprise calculating 306 at least a first set of respective approximations based at least in part on a subset of the input data array, one or more approximation functions (*e.g.*, based on a least squares approach or a polynomial function). The instructions may comprise re-calculating 308 at least a second set of respective approximations based at least in part on one or more different subsets of the input data array, one or more approximation functions. As stated before, these calculations may be based at least in part on the aforementioned formulae. At 310, at least one of the first or second sets of approximations are stored and corresponding confidence interval in the output data array for at least one of the elements based at least in part on selecting the approximation with a lower corresponding confidence interval. Further, at 310, one or more approximations may be rejected based at least in part on a statistical significance of the residual sum of squares (RSS):

$RSS_i \geq \sigma^2 \cdot \chi^2(\alpha, n-p)$, where:

(n-p) represents a number of degrees of freedom;

p represents a number of functions used for approximation;

n represents an approximation window;

α = (1 - confidence level);

in a case of a priori known standard deviation of the noise of the signal.

[0058] In a first such embodiment, the exemplary method 300 may be implemented as a set of instructions stored on a memory component (such as a memory circuit, a platter of a hard disk drive, a solid-state storage device, or a magnetic or optical disc) of a device, where such instructions are configured to, when executed on a processor of the device, perform the elements of the techniques presented herein. In a second such embodiment, the exemplary method 300 may be implemented as a set of hardware components comprising a system, such as logic gates of an electronic circuit or components of a software architecture, that interoperate to perform the elements of the techniques presented herein. Additional embodiments may also be devised, such as solutions that utilize hybrid components (*e.g.*, a field-programmable gate array (FPGA) programmed to implement the elements of the exemplary method 300) and solutions utilizing a combination of hardware and software elements. Many variations in such embodiments may be devised that implement the techniques presented herein, such as the elements of the exemplary method 300 of Fig. 3.

[0059] A smoothing method may comprise an input signal array (*e.g.*, input data set), each element of the array having two coordinates: X (independent coordinate) and Y (result of measurement, dependent coordinate); an output signal array with at least one element, each element having two coordinates (X and Y similar to input array) and a confidence value associated with every element of the output array, X coordinates (target positions) of the input array are defined in advance by the nature of the approximation task; and a set of one or more procedures capable of producing an output signal value for a target position and its associated confidence interval value from the subset of values in said input signal array. Multiple (*e.g.*, at least two) approximations using different functions or different input data subsets should be made and the correct approximation is selected using comparison of confidence intervals.

[0060] Figure 7 illustrates an example where original raw data is graphed using a dotted line, while SG filtering is shown as the thinner solid line, and an adaptive confidence interval non-central approximation filter is shown with the thicker solid line. The filter gap is the same in both the SG and confidence interval case and equals 41 points. The use of the SG filter with a reasonable gap produces artifacts which are difficult to handle. Further, the SG filter distorts peaks near steps and cannot be used to process the signal correctly. At the second peak of Figure 7, the distortion becomes especially pronounced. In contrast to the SG filter, the confidence interval filtering with non-central approximations produces no significant distortions while achieving noise suppression in an effective manner. Therefore, confidence interval noise filters with non-central approximations may provide more accurate results, especially for data comprising sudden, abrupt changes of signal levels.

Filtering Based On Variable Window And/Or Degree Of Polynomial

[0061] Adjustments may be made to the window size and/or the degree of the polynomial used for modeling the confidence interval. Accordingly, a window size may be based on one or more locations of signal characteristics. In one embodiment, a maximum window size may be selected to avoid hiding small objects. Smaller windows may provide better estimates for steep slopes, steps, and/or jumps, while bigger windows may provide better noise reduction levels for lengthy baseline regions. Additionally, optimal window width may depend on the point neighborhood. Therefore, the larger the distance to the nearest peak, the wider the approximation window can be. Fit quality criteria based on noise estimates may also be applied to avoid effects of accidental good fits for small approximation windows and peak suppression for wider windows. As an example, noise filtering may be implemented with a variable window at a fixed degree (*e.g.*, cubic, $p = 3$), where $\delta = 0.025$ and logarithmic steps are used, increasing or decreasing window $\sqrt{2}$ times with each step. Filtering may be performed with the noise definition window (NDW) with a maximal gap G_0. In the following example, three steps are performed upwards and three downwards, increasing overall window width range to 8:

$$\frac{G_0}{2\sqrt{2}}, \frac{G_0}{2}, \frac{G_0}{\sqrt{2}}, G_0, G_0 \cdot \sqrt{2}, G_0 \cdot 2, G_0 \cdot 2\sqrt{2}$$

[0062] Figure 8 illustrates two peaks of different heights to simulate chromatographic analysis. The model peaks have a shape of exponentially-modified Gaussian function:

$$F(t) = \frac{h \cdot \sigma}{\tau}\sqrt{\frac{\pi}{2}} \cdot e^{\left(\frac{T-t}{\tau} + \frac{\sigma^2}{2\tau^2}\right)} \cdot \operatorname{erfc}\left(\left(\frac{T-t}{\sigma} + \frac{\sigma}{\tau}\right)\middle/\sqrt{2}\right)$$

t is a time or position coordinate

T is a parameter which defines the position of the peak

σ and τ are parameters which define width and asymmetry of the peak, respectively. For this example, a normally distributed noise produced with pseudo-random noise generator may be added to the signal and parameters of the peaks may be selected so that $\frac{\tau}{\sigma} = 3$ and the left slope has 11 sampling points from half-height level to the peak top. Assuming the first peak is at the detection limit and the second peak is at the quantification limit, then:

$$\frac{2H_1}{h_N} = (S / N) = 3$$

$$\frac{2H_2}{h_N} = (S / N) = 10$$

where H_1 and H_2 are height of the first and second peaks, respectively;

$h_N \approx 6\sigma_N$ - range of background at baseline; and

σ_N represents the variance of normally distributed noise.

[0063] As noted, conventional SG filtering in the case of aggressive noise suppression tends to distort the peak. The higher the peak, the bigger the distortion, which is measured by the maximum of difference between the smoothed and the initial signal. Because of this, a balance between distortion and noise suppression by varying of the gap of the filter may be found. However, with an artificial signal this is a relatively simple task due to the fact that the precise original shape of the peaks is known. The gap of the Savitzky-Golay filter may be selected so that distortion for the second, higher peak is approximately twice as large as the background noise range at the baseline. For this sampling rate, the gap selected by this condition appeared to be equal to 31. Thus, the maximal G_0 for the adaptive filter which produces approximately the same maximal distortion near the peak top may be found. A series of reference signals with different noise samples may be analyzed and when G_0 is about 161, the distortion of the highest peak is approximately the same as for previously used conventional SG filter, as illustrated by Figure 9. Signal filtering with a conventional SG filter at a gap of 31 (SG31) and an adaptive gap selection with non-central approximation with base gap G_0 of 161 (AGSG160). The pure signal is drawn for comparison purposes.

[0064] Figure 10 illustrates an example which makes the comparison between the filtered and pure signals more apparent. Both the SG and confidence filters produce approximately the same distortion for the second (higher) peak at the quantification limit. For lower peaks, the adaptive confidence filter may produce bigger distortions than the conventional Savitzky-Golay filter when large G_0 is used, although these distortions are much smaller than those that would be caused by SG with gap G_0. The difference (*e.g.*, signal distortion) between the filtered and the pure signal may be illustrated by Figure 10 where dSG31 (dotted line) represents filtering with conventional SG at a gap of 31, while dAGSG (solid line) represents an adaptive confidence gap selection with non-central approximation with base gap of 161.

[0065] It should be noted that the distortion varies very little as G_0 varies from 141 to 241. This is natural since, in all cases, the adaptive algorithm chooses an appropriate filtering gap for the top of the peak, as shown at Figure 11. Meanwhile, the maximum gap is selected for noise suppression at the flat baseline piece of the signal. This produces the appropriate noise suppression. Gap values are selected based on best approximations (*e.g.*, lowest confidence intervals).

[0066] Figure 12 illustrates noise suppression at baseline for conventional SG (using a gap of 31, shown as a solid line) and for the confidence filtering with an adaptive gap (non-central approximation with base gap of $G_0 = 161$, thick line). The original noise is shown for comparison as a dotted line.

The noise suppression generated by conventional Savitzky-Golay filter is:

$$R_{SG} = \frac{h_0}{h_{SG}} \approx 5$$

For the adaptive filter presented herein:

$$R_a = \frac{h_0}{h_a} \approx 13$$

Therefore, noise suppression of the adaptive filter is almost three times better than that of the conventional SG filter with the same distortion for the second peak.

[0067] Figure 13 illustrates a series of peaks with increasing heights for distortion investigation, while Figure 14 illustrates the difference (*e.g.*, signal distortion) between a filtered signal and a pure signal from Figure 13. Where dSG31 represents filtering with conventional Savitzky-Golay at gap 31 (curve lowered with respect to 0) and dASG160 is the confidence filter with adaptive gap selection and non-central approximation, largest (base) gap of $G_0 = 161$.

[0068] One of the benefits of the confidence filter with an adaptive slit is that the absolute value of the peak distortion does not have a strong dependence on the peak height, as shown in Figures 13 and 14. Therefore, minimal statistical confidence intervals are defined by noise level rather than peak shape or peak height. The bigger the peak, the less the relative peak distortion because absolute distortion is nearly constant. Conventional SG filtering fails to provide this since its distortion directly relates to peak height. The peak at detection limit may be considered to be the worst case for the confidence filter; distortion being the price for significant improvement of baseline noise due to the bigger gap.

White Noise

[0069] White homoscedastic noise is generally the simplest case for most filters. Figure 15A demonstrates the confidence interval filter applied to a noisy signal to smooth an artificial chomatogram of an electromyography (EMG) peak with white noise. The original data is represented by the solid line, while the dashed line represents the signal after the confidence interval filter has been applied. The noise definition width for Figure 15A is 31. Figure 15B illustrates a graph of the approximation window which corresponds to the respective window position of Figure 15A. Baseline regions of Figure 15A are approximated with wider windows (*e.g.*, in the 30-40 range) than the peak region. Since the approximation windows along the baseline are wider, this results in better noise reduction. A comparison between SG filtering and confidence interval filtering can be made by evaluating the minimal and maximal window widths within a smoothing region. In order to maintain the shape of a peak (*e.g.*, keep the peak shape unchanged), the data should be smoothed with minimal window widths and additional noise reduction may be evaluated as follows:

$$noise_{SG} / noise_{confidence} \approx \sqrt{\frac{(n-p)_{\max}}{(n-p)_{\min}}}$$

and in the case of algorithm applied and data from Figure 15A can be as high as $8^{1/2} \approx 2.8$.

Non-White Noise

[0070] Generally, non-white noise may be trickier to deal with in comparison to white noise, and results may depend on the particular noise model being used. As an example, pump pulsations may frequently occur in chromatography. Figure 16 illustrates pump pulsations suppressed by a confidence interval filter using multiple noise definition windows. The original data of Figure 16 is shown on the legend as “Noisy”, while the dashed line illustrates the confidence filter with a width NDW = 121, and “Conf10” illustrates a confidence filter with NDW = 11. It will be appreciated that the curves have been vertically shifted for ease of comparison and to prevent overlapping. Various values may be chosen for NDW, and these may based on the requirements for filtering, the purpose of the research, and/or the aim of application of the filter. If the point of applying the filter is to reveal peaks, selection of a higher noise definition window may be appropriate. As illustrated in Figure 16, the time of the pump stroke is smaller than the peak width, therefore, the confidence filter provides results accordingly. Alternatively, if the purpose of the filter is to study the shape of minute pulsations, a small expected noise level or a narrow window (*e.g.*, NDW = 11) may smooth pulsations and aid construction of a better pump pulsation model. Further, noise reduction for regular or periodic noise may be better than for white noise alone, depending on particular features of the noise.

Confidence Filtering For Capillary Electrophoresis

[0071] Confidence filtering may be applied to capillary electrophoresis (CE), as some CE peaks are extremely narrow, while other may be triangular. Figure 17A illustrates an example CE signal (shown in the middle as a solid line, labeled “Capel”), characterized by the triangular peak, a SG filter applied to the CE signal with a width = 35 (“SG35”), and the confidence interval filter with a NDW = 101 applied as a dotted line (“Conf101”). Figure 17B illustrates the polynomial width (shown as a dashed line) versus the distance of the point used for approximation from the center of the polynomial (shown as the solid line). Figure 17C illustrates the confidence interval of the approximations by the confidence filter. The confidence filter produced an extremely clean result, as the baseline of the CE signal and the SG filtered signal are both slightly sinusoidal between 1400 and 1500. The confidence interval filter, on the other hand, flattens the baseline into a desirable line shape and eliminates much of the unwanted noise while creating less of a disturbance near the steep end of the peak. It will be appreciated that the approximations near the top of the peak are based on non-central approximations. Additionally, the

SG approximation corresponds to the smallest window of the implementation of the confidence interval filter. Most filters cannot be applied to such data, so the improvement in noise reduction may be evaluated as follows:

$$noise_{before} / noise_{after} \approx \sqrt{(n-p)_{\max}}$$

and can come up to 10…20 depending on data rate and particular data.

Outliers

[0072] According to one aspect, the confidence interval filter does not reject outliers, but simply avoids them. For example, a modified procedure based on robust regression methods may eliminate outliers. However, drawbacks to such an approach may exist, as robust regressions may obscure the errors of selecting an improper approximation model. Therefore, it may be advantageous to perform object-dependent outlier elimination separately. Alternatively, outlier elimination may be based on identification of outliers as points which were not estimated or points with the highest confidence intervals. These points may be re-evaluated and select points from the input data set may be excluded.

[0073] Still another embodiment involves a computer-readable medium comprising processor-executable instructions configured to implement one or more of the techniques presented herein. An exemplary computer-readable medium that may be devised in these ways is illustrated in Figure 18, wherein the implementation 2000 comprises a computer-readable medium 2008 (*e.g.*, a CD-R, DVD-R, or a platter of a hard disk drive), on which is encoded computer-readable data 2006. This computer-readable data 2006 in turn comprises a set of computer instructions 2004 configured to operate according to one or more of the principles set forth herein. In one such embodiment 2002, the processor-executable instructions 2004 may be configured to perform a method, such as at least some of the exemplary method 100 of Figure 1, for example. In another such embodiment, the processor-executable instructions 2004 may be configured to implement a system, such as at least some of an exemplary system, for example. Many such computer-readable media may be devised by those of ordinary skill in the art that are configured to operate in accordance with the techniques presented herein.

[0074] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

[0075] As used in this application, the terms "component", "module", "system", "interface", and the like are generally intended to refer to a computer-related entity, either hardware, a combination of hardware and software, software, or software in execution. For example, a component may be, but is not limited to being, a process running on a processor, a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a controller and the controller can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers.

[0076] Furthermore, the claimed subject matter may be implemented as a method, apparatus, or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof to control a computer to implement the disclosed subject matter. The term "article of manufacture" as used herein is intended to encompass a computer program accessible from any computer-readable device, carrier, or media. Of course, those skilled in the art may recognize many modifications may be made to this configuration without departing from the scope or spirit of the claimed subject matter.

[0077] Figure 19 and the following discussion provide a brief, general description of a suitable computing environment to implement embodiments of one or more of the provisions set forth herein. The operating environment of Figure 19 is only one example of a suitable operating environment and is not intended to suggest any limitation as to the scope of use or functionality of the operating environment. Example computing devices include, but are not limited to, personal computers, server computers, hand-held or laptop devices, mobile devices (such as mobile phones, Personal Digital Assistants

(PDAs), media players, and the like), multiprocessor systems, consumer electronics, mini computers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

[0078] Although not required, embodiments are described in the general context of "computer readable instructions" being executed by one or more computing devices. Computer readable instructions may be distributed via computer readable media (discussed below). Computer readable instructions may be implemented as program modules, such as functions, objects, Application Programming Interfaces (APIs), data structures, and the like, that perform particular tasks or implement particular abstract data types. Typically, the functionality of the computer readable instructions may be combined or distributed as desired in various environments.

[0079] Figure 19 illustrates an example of a system 2100 comprising a computing device 2112 configured to implement one or more embodiments provided herein. In one configuration, computing device 2112 includes at least one processing unit 2116 and memory 2118. Depending on the exact configuration and type of computing device, memory 2118 may be volatile (such as RAM, for example), non-volatile (such as ROM, flash memory, etc., for example), or some combination of the two. This configuration is illustrated in Figure 19 by dashed line 2114.

[0080] In other embodiments, device 2112 may include additional features and/or functionality. For example, device 2112 may also include additional storage (*e.g.*, removable and/or non-removable) including, but not limited to, magnetic storage, optical storage, and the like. Such additional storage is illustrated in Figure 19 by storage 2120. In one embodiment, computer readable instructions to implement one or more embodiments provided herein may be in storage 2120. Storage 2120 may also store other computer readable instructions to implement an operating system, an application program, and the like. Computer readable instructions may be loaded in memory 2118 for execution by processing unit 2116, for example.

[0081] The term "computer readable media" as used herein includes computer storage media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions or other data. Memory 2118 and storage 2120 are examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, Digital Versatile Disks (DVDs) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by device 2112. Any such computer storage media may be part of device 2112.

[0082] The term “computer readable media” may include communication media. Communication media typically embodies computer readable instructions or other data in a “modulated data signal” such as a carrier wave or other transport mechanism and includes any information delivery media. The term “modulated data signal” may include a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal.

[0083] Device 2112 may also include communication connection(s) 2126 that allows device 2112 to communicate with other devices. Communication connection(s) 2126 may include, but is not limited to, a modem, a Network Interface Card (NIC), an integrated network interface, a radio frequency transmitter/receiver, an infrared port, a USB connection, or other interfaces for connecting computing device 2112 to other computing devices. Communication connection(s) 2126 may include a wired connection or a wireless connection. Communication connection(s) 2126 may transmit and/or receive communication media.

[0084] Device 2112 may include input device(s) 2124 such as keyboard, mouse, pen, voice input device, touch input device, infrared cameras, video input devices, and/or any other input device. Output device(s) 2122 such as one or more displays, speakers, printers, and/or any other output device may also be included in device 2112. Input device(s) 2124 and output device(s) 2122 may be connected to device 2112 via a wired connection, wireless connection, or any combination thereof. In one embodiment, an input device or an output device from another computing device may be used as input device(s) 2124 or output device(s) 2122 for computing device 2112.

[0085] Components of computing device 2112 may be connected by various interconnects, such as a bus. Such interconnects may include a Peripheral Component Interconnect (PCI), such as PCI Express, a Universal Serial Bus (USB), firewire (IEEE 1394), an optical bus structure, and the like. In another embodiment, components of computing device 2112 may be interconnected by a network. For example, memory 2118 may be comprised of multiple physical memory units located in different physical locations interconnected by a network.

[0086] Those skilled in the art may realize that storage devices utilized to store computer readable instructions may be distributed across a network. For example, a computing device 2130 accessible via network 2128 may store computer readable instructions to implement one or more embodiments provided herein. Computing device 2112 may access computing device 2130 and download a part or all of the computer readable instructions for execution. Alternatively, computing device 2112 may download pieces of the computer readable instructions, as needed, or some instructions may be executed at computing device 2112 and some at computing device 2130.

[0087] Various operations of embodiments are provided herein. In one embodiment, one or more of the operations described may constitute computer readable instructions stored on one or more computer readable media, which if executed by a computing device, will cause the computing device to perform the operations described. The order in which some or all of the operations are described should not be construed as to imply that these operations are necessarily order dependent. Alternative ordering will be appreciated by one skilled in the art having the benefit of this description. Further, it will be understood that not all operations are necessarily present in each embodiment provided herein.

[0088] Moreover, the word "exemplary" is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as "exemplary" is not necessarily to be construed as advantageous over other aspects or designs. Rather, use of the word exemplary is intended to present concepts in a concrete fashion. As used in this application, the term "or" is intended to mean an inclusive "or" rather than an exclusive "or". That is, unless specified otherwise, or clear from context, "X employs A or B" is intended to mean any of the natural inclusive permutations. That is, if X employs A; X employs B; or X employs both A and B, then "X employs A or B" is satisfied under any of the foregoing instances. In addition, the articles "a" and "an" as used in this application and the appended claims may generally be construed to mean "one or more" unless specified otherwise or clear from context to be directed to a singular form. Further, at least one of A and B and/or the like generally means A or B, or both A and B.

[0089] Although the disclosure is shown and described with respect to one or more implementations, equivalent alterations and modifications will occur to others skilled in the art based upon a reading and understanding of this specification and the annexed drawings. The disclosure includes all such modifications and alterations and is limited only by the scope of the following claims. In particular regard to the various functions performed by the above described components (*e.g.*, elements, resources, etc.), the terms used to describe such components are intended to correspond, unless otherwise indicated, to any component which performs the specified function of the described component (*e.g.,* that is functionally equivalent), even though not structurally equivalent to the disclosed structure which performs the function in the herein illustrated exemplary implementations of the disclosure. In addition, while a particular feature of the disclosure may have been disclosed with respect to only one of several implementations, such feature may be combined with one or more other features of the other implementations as may be desired and advantageous for any given or particular application. Furthermore, to the extent that the terms "includes", "having", "has", "with", or variants thereof are used in either the detailed description or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising."

CLAIMS

What is claimed is:

1. A method (300) for filtering noise from a signal using a computer having a processor, the method (300) comprising:

executing on the processor instructions configured to:

receive (302) a signal comprising noise and store the signal in an input data array of three or more signal values which correspond to a set of three or more points, the points comprising an independent X coordinate and a measured Y coordinate;

an output data array (304) comprising a number of elements based at least in part on a selected number of targets to approximate, the respective elements comprising an approximation and a confidence interval associated with the respective approximation, the approximation comprising an X coordinate and a Y coordinate, the number of targets defined by a nature of a task;

calculate (306) at least a first set of respective approximations based at least in part on a subset of the input data array, one or more approximation functions;

re-calculate (308) at least a second set of respective approximations based at least in part on one or more different subsets of the input data array, one or more approximation functions; and

store (310) at least one of the first or second sets of approximations and corresponding confidence interval in the output data array for at least one of the elements based at least in part on selecting the approximation with a lower corresponding confidence interval.

2. The method of claim 1, at least one of the approximation functions based at least in part on a least squares approach.

3. The method of claim 1, at least one of the approximation functions comprising one or more polynomial functions.

4. The method of claim 1, comprising scaling the signal based at least in part on at least one of a scaling function and an inverse scaling function.

5. The method of claim 1, comprising sorting the input data array in an ascending order based at least in part on at least one of the X coordinates.

6. The method of claim 5, comprising selecting at least one of the subsets of the input data array based at least in part on a defined number (approximation window, *n*) of adjacent input points.

7. The method of claim 2, comprising rejecting of at least one of the approximations by one or more criterion based at least in part on a statistical significance of the residual sum of squares (RSS)

$RSS_i \geq \sigma^2 \cdot \chi^2(\alpha, n - p)$, where:

(n-p) represents a number of degrees of freedom;

p represents a number of functions used for approximation;

n represents an approximation window;

α = (1 - confidence level);

in a case of a priori known standard deviation of the noise of the signal.

8. The method of claim 7, comprising a confidence level based at least in part on a user-defined parameter of the algorithm.

9. The method of claim 2, calculating or re-calculating the confidence intervals based at least in part on:

$$C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$$; where:

C_Y represents the confidence interval;

$\mathbf{Y} = \{ y_1, y_2 \ldots, y_n \}$ represents the set of points for the signal;

t represents the Student's coefficient;

α = (1 - confidence level);

$n - p$ represents a number of degrees of freedom;

$$S^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p};$$

$\hat{\boldsymbol{\beta}}$ represents one or more polynomial coefficients for regression;

X represents a matrix of approximating function :

$$\mathbf{X} = \begin{Bmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) & \ldots & f_p(x_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1(x_n) & f_2(x_n) & f_3(x_n) & \ldots & f_p(x_n) \end{Bmatrix}$$

$$u_* = \mathbf{x}'_* (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{x}_* ;$$

x_i represents the independent variable from the corresponding elements of the input data array;

$\mathbf{x}'_* = \{ f_1(x_*), f_2(x_*), \ldots, f_p(x_*) \}$; and

x_* represents a position of the approximated value.

10. The method of claim 2, calculating or re-calculating the confidence intervals based at least in part on:

$$C_Y = t_{n-p}^{\frac{1}{2}\alpha} \cdot \sigma \cdot \sqrt{u_*}$$; where:

C_Y represents the confidence interval;

$\mathbf{Y} = \{ y_1, y_2 \ldots, y_n \}$ represents the set of points for the signal;

$\hat{\mathbf{Y}} = \{ \hat{y}_1, \hat{y}_2 \ldots, \hat{y}_n \}$ represents the set of approximated values;

$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$;

$\hat{\boldsymbol{\beta}}$ represents one or more polynomial coefficients for regression;

t represents the student's coefficient for a confidence level;

α = (1 - confidence level);

n - p represents a number of degrees of freedom;

$\hat{\boldsymbol{\beta}}$ represents one or more polynomial coefficients for regression;

X represents a matrix of approximating function:

$$\mathbf{X} = \begin{Bmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) & \ldots & f_p(x_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1(x_n) & f_2(x_n) & f_3(x_n) & \ldots & f_p(x_n) \end{Bmatrix}$$

$$u_* = \mathbf{x}_*'(\mathbf{X}' \cdot \mathbf{X})^{-1}\mathbf{x}_*$$;

x_i represents the independent variable from the corresponding elements of the input data array;

$$\mathbf{x}_*' = \{ f_1(x_*), f_2(x_*), \ldots, f_p(x_*) \}$$;

x_* represents a position of the approximated value; and

σ^2 represents a noise variance.

11. The method of claim 2, calculating or re-calculating the confidence intervals based at least in part on:

$$C_Y = \begin{Bmatrix} S^2 < \sigma^2, Formula\,(a) \\ S^2 \geq \sigma^2, Formula\,(b) \end{Bmatrix}$$; where:

Formula (a) $C_Y = t_{n-p}^{½\alpha} \cdot \sigma \cdot \sqrt{u_*}$;

Formula (b) $C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$;

C_Y represents the confidence interval;

$\mathbf{Y} = \{ y_1, y_2 \ldots, y_n \}$ represents the set of points for the signal;

$\hat{\mathbf{Y}} = \{ \hat{y}_1, \hat{y}_2 \ldots, \hat{y}_n \}$ represents the set of approximated values;

$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$;

t represents the student's coefficient for a confidence level;

α = (1 - confidence level);

$n - p$ represents a number of degrees of freedom;

$$S^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p};$$

$\hat{\boldsymbol{\beta}}$ represents one or more polynomial coefficients for regression;

X represents a matrix of approximating function:

$$\mathbf{X} = \begin{Bmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) & \ldots & f_p(x_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1(x_n) & f_2(x_n) & f_3(x_n) & \ldots & f_p(x_n) \end{Bmatrix}$$

$u_* = \mathbf{x}'_* (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{x}_*$;

x_i represents the independent variable from the corresponding elements of the input data array;

$\mathbf{x}'_* = \{ f_1(x_*), f_2(x_*), \ldots, f_p(x_*) \}$;

x_* represents a position of the approximated value; and

σ^2 represents a noise variance.

12. A computer-readable storage device (2008) comprising computer-executable instructions (2004), which when executed at least in part via a processor on a computer perform acts, comprising:

receiving (302) a signal comprising noise and an input data array of three or more signal values which correspond to a set of three or more points, the points comprising an independent X coordinate and a measured Y coordinate;

approximating (304) an output data array comprising a number of elements based at least in part on a selected number of targets to approximate, the respective elements comprising an approximation and a confidence interval associated with the respective approximation, the approximation comprising an X coordinate and a Y coordinate, the number of targets defined by a nature of a task;

calculating (306) at least a first set of respective approximations based at least in part on a subset of the input data array, one or more approximation functions;

re-calculating (308) at least a second set of respective approximations based at least in part on one or more different subsets of the input data array, one or more approximation functions; and

storing (310) at least one of the first or second sets of approximations and corresponding confidence interval in the output data array for at least one of the elements based at least in part on selecting the approximation with a lower corresponding confidence interval.

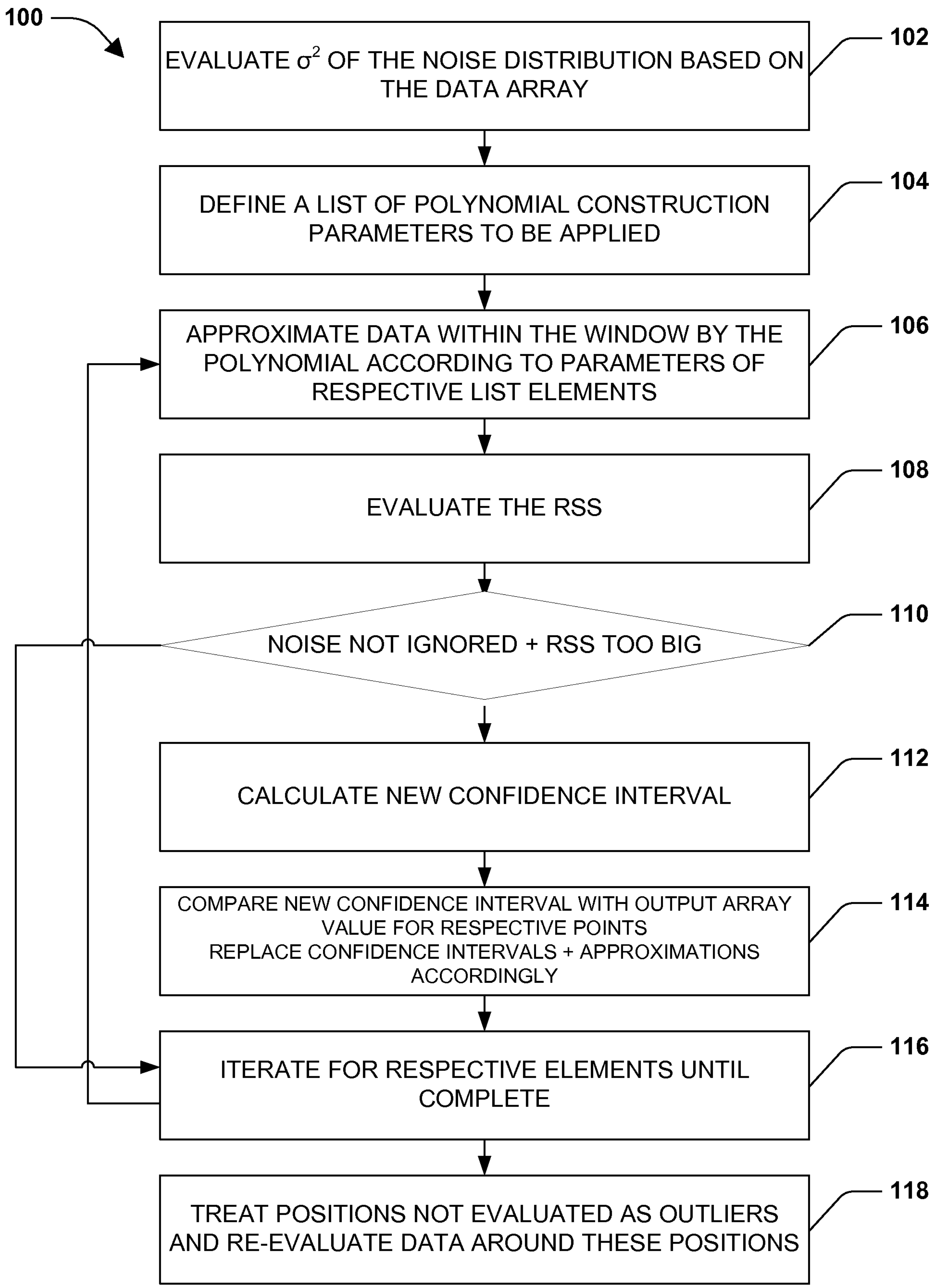
100
EVALUATE σ² OF THE NOISE DISTRIBUTION BASED ON THE DATA ARRAY
102
DEFINE A LIST OF POLYNOMIAL CONSTRUCTION PARAMETERS TO BE APPLIED
104
APPROXIMATE DATA WITHIN THE WINDOW BY THE POLYNOMIAL ACCORDING TO PARAMETERS OF RESPECTIVE LIST ELEMENTS
106
EVALUATE THE RSS
108
NOISE NOT IGNORED + RSS TOO BIG
110
CALCULATE NEW CONFIDENCE INTERVAL
112
COMPARE NEW CONFIDENCE INTERVAL WITH OUTPUT ARRAY VALUE FOR RESPECTIVE POINTS
REPLACE CONFIDENCE INTERVALS + APPROXIMATIONS ACCORDINGLY
114
ITERATE FOR RESPECTIVE ELEMENTS UNTIL COMPLETE
116
TREAT POSITIONS NOT EVALUATED AS OUTLIERS AND RE-EVALUATE DATA AROUND THESE POSITIONS
118

FIG. 1

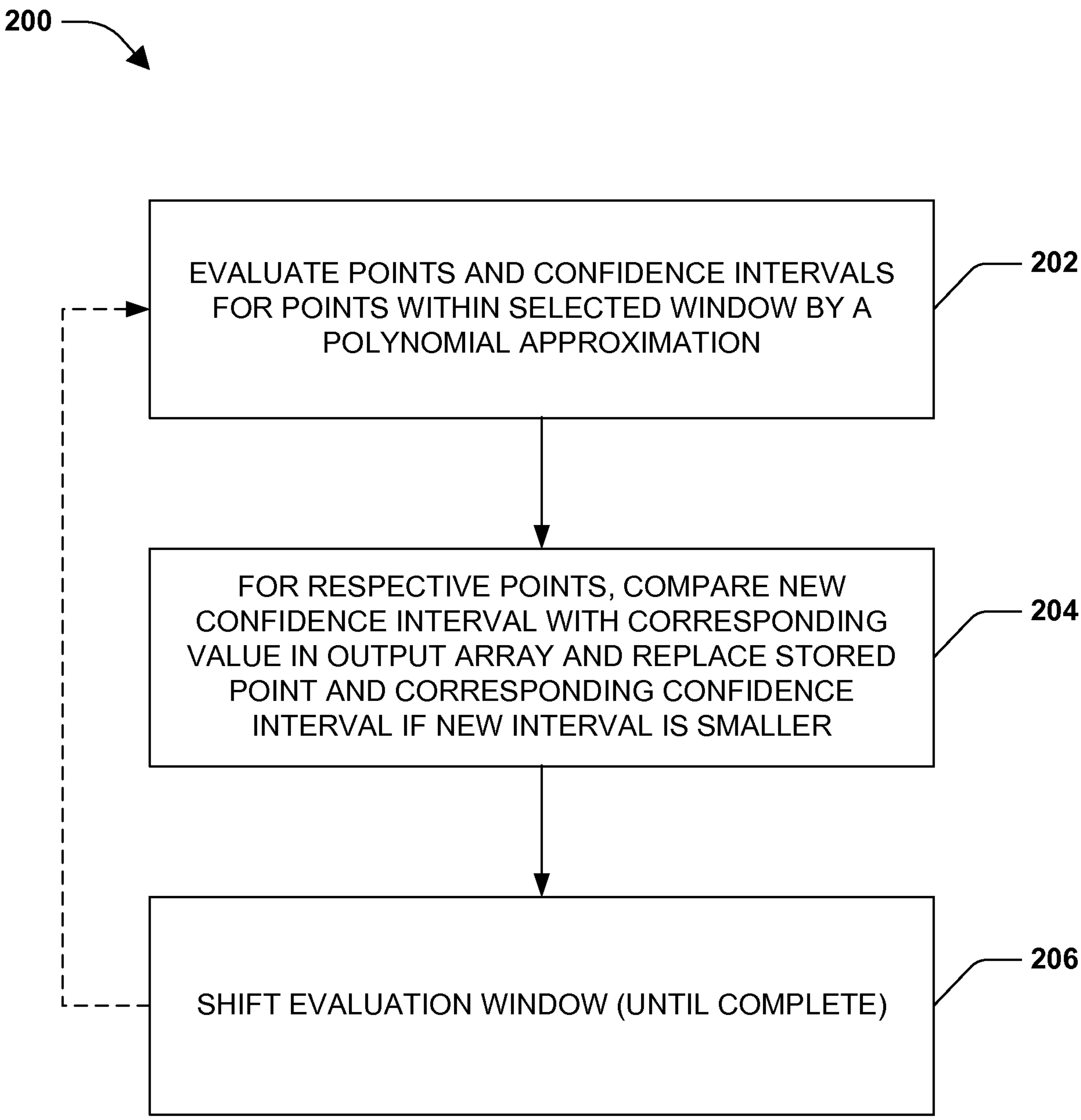
200
EVALUATE POINTS AND CONFIDENCE INTERVALS FOR POINTS WITHIN SELECTED WINDOW BY A POLYNOMIAL APPROXIMATION
202
FOR RESPECTIVE POINTS, COMPARE NEW CONFIDENCE INTERVAL WITH CORRESPONDING VALUE IN OUTPUT ARRAY AND REPLACE STORED POINT AND CORRESPONDING CONFIDENCE INTERVAL IF NEW INTERVAL IS SMALLER
204
SHIFT EVALUATION WINDOW (UNTIL COMPLETE)
206

FIG. 2

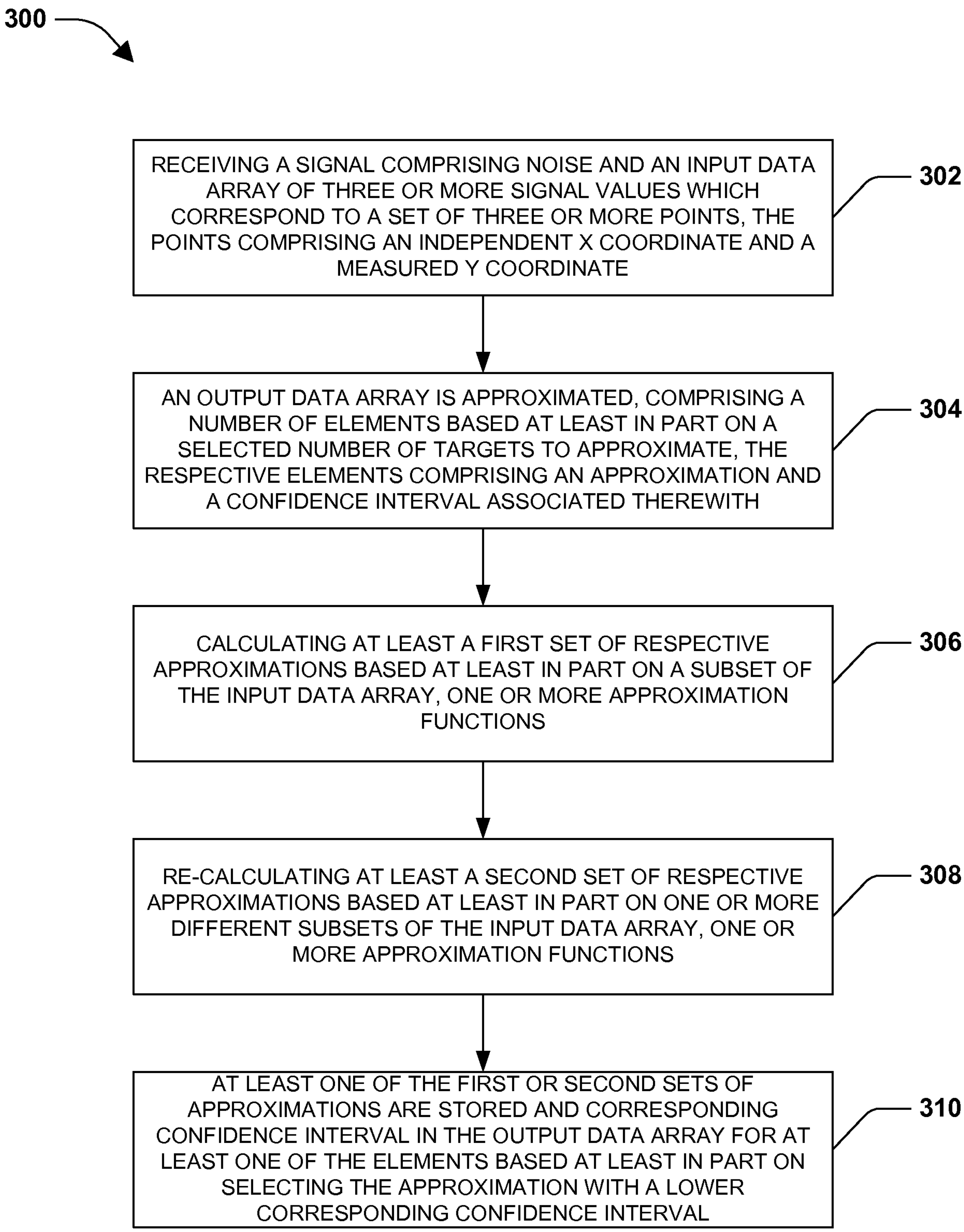
300
RECEIVING A SIGNAL COMPRISING NOISE AND AN INPUT DATA ARRAY OF THREE OR MORE SIGNAL VALUES WHICH CORRESPOND TO A SET OF THREE OR MORE POINTS, THE POINTS COMPRISING AN INDEPENDENT X COORDINATE AND A MEASURED Y COORDINATE
302
AN OUTPUT DATA ARRAY IS APPROXIMATED, COMPRISING A NUMBER OF ELEMENTS BASED AT LEAST IN PART ON A SELECTED NUMBER OF TARGETS TO APPROXIMATE, THE RESPECTIVE ELEMENTS COMPRISING AN APPROXIMATION AND A CONFIDENCE INTERVAL ASSOCIATED THEREWITH
304
CALCULATING AT LEAST A FIRST SET OF RESPECTIVE APPROXIMATIONS BASED AT LEAST IN PART ON A SUBSET OF THE INPUT DATA ARRAY, ONE OR MORE APPROXIMATION FUNCTIONS
306
RE-CALCULATING AT LEAST A SECOND SET OF RESPECTIVE APPROXIMATIONS BASED AT LEAST IN PART ON ONE OR MORE DIFFERENT SUBSETS OF THE INPUT DATA ARRAY, ONE OR MORE APPROXIMATION FUNCTIONS
308
AT LEAST ONE OF THE FIRST OR SECOND SETS OF APPROXIMATIONS ARE STORED AND CORRESPONDING CONFIDENCE INTERVAL IN THE OUTPUT DATA ARRAY FOR AT LEAST ONE OF THE ELEMENTS BASED AT LEAST IN PART ON SELECTING THE APPROXIMATION WITH A LOWER CORRESPONDING CONFIDENCE INTERVAL
310

FIG. 3

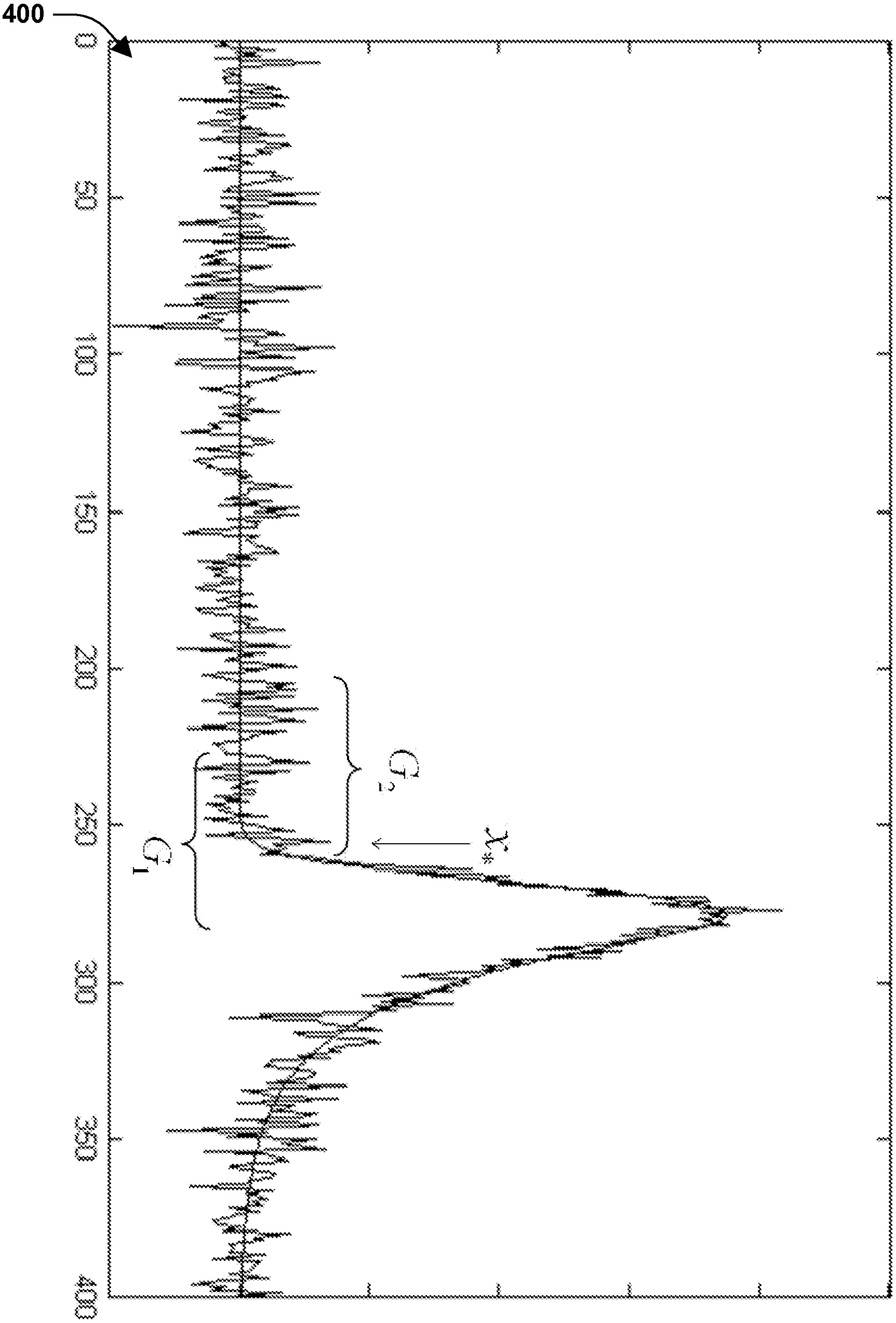

FIG. 4

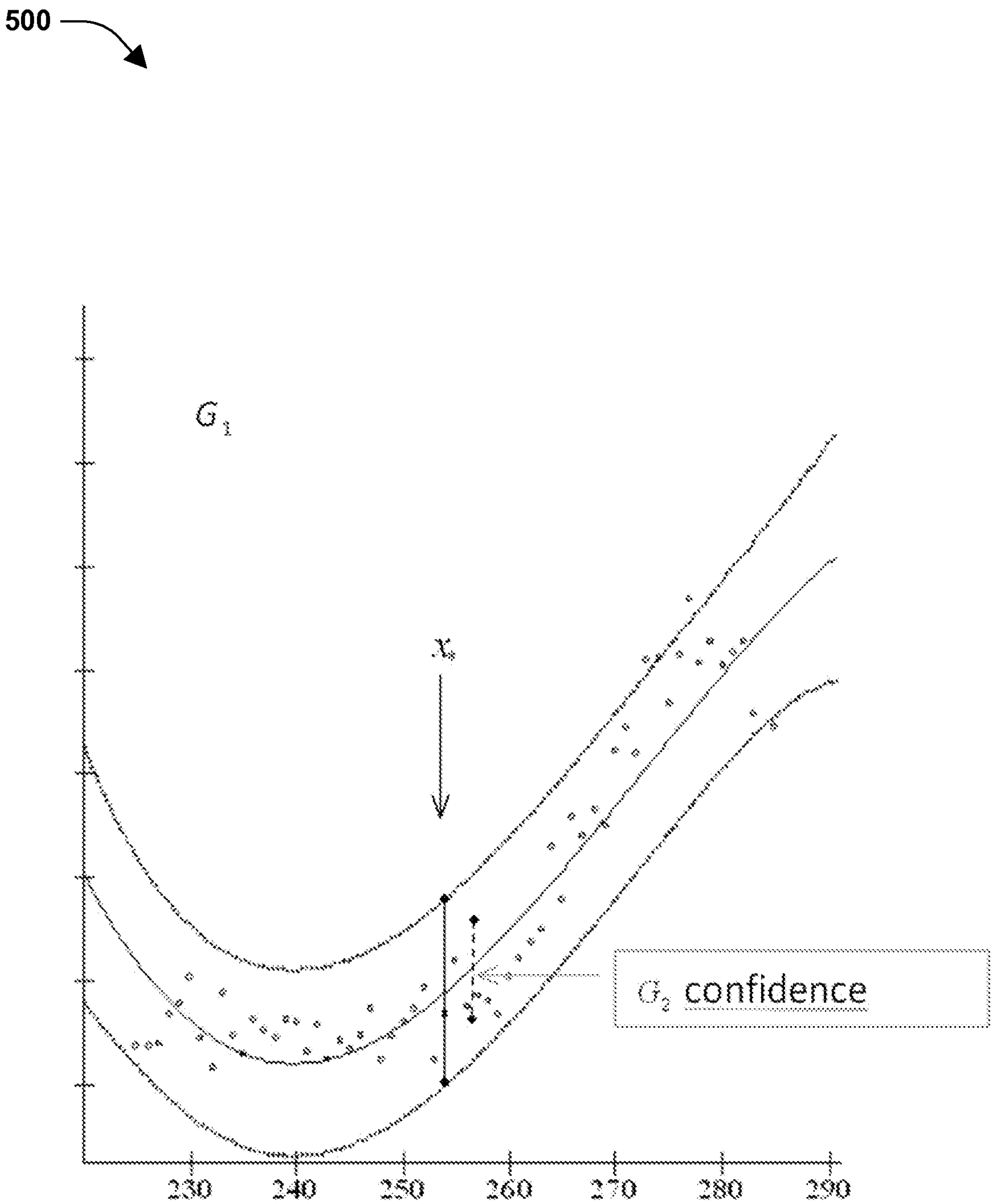

FIG. 5

600

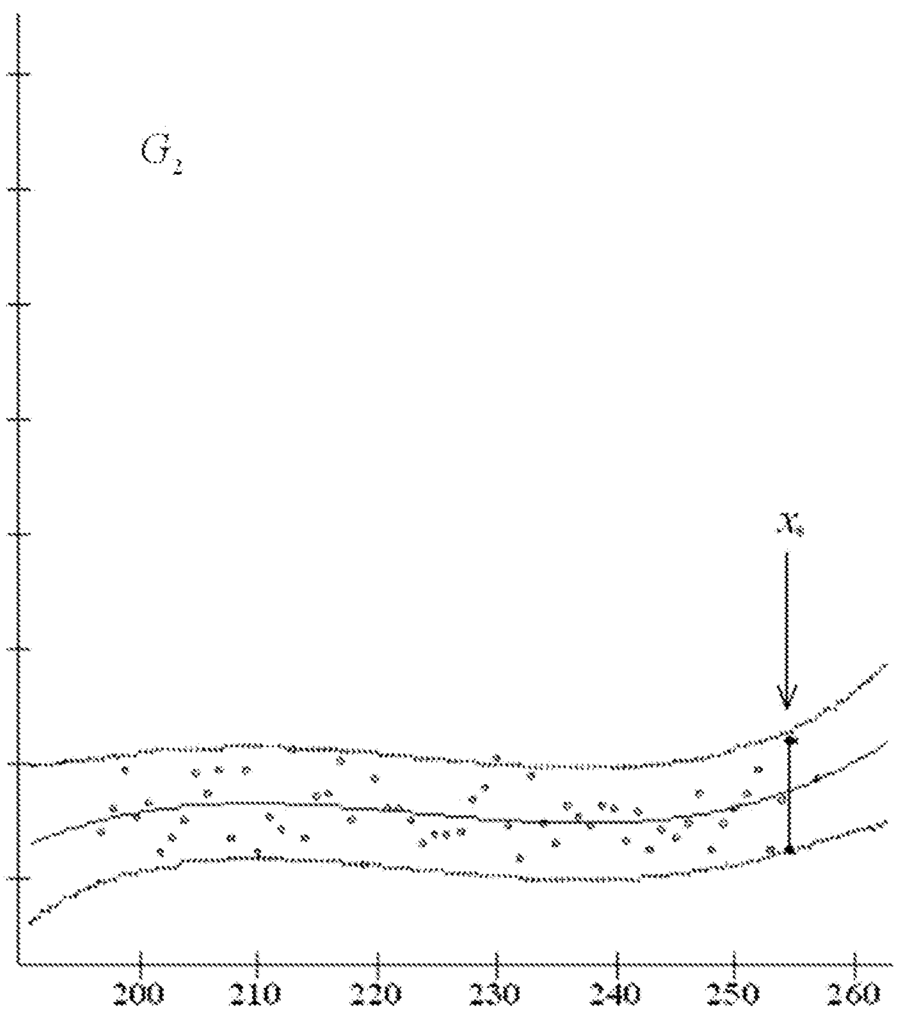

FIG. 6

700

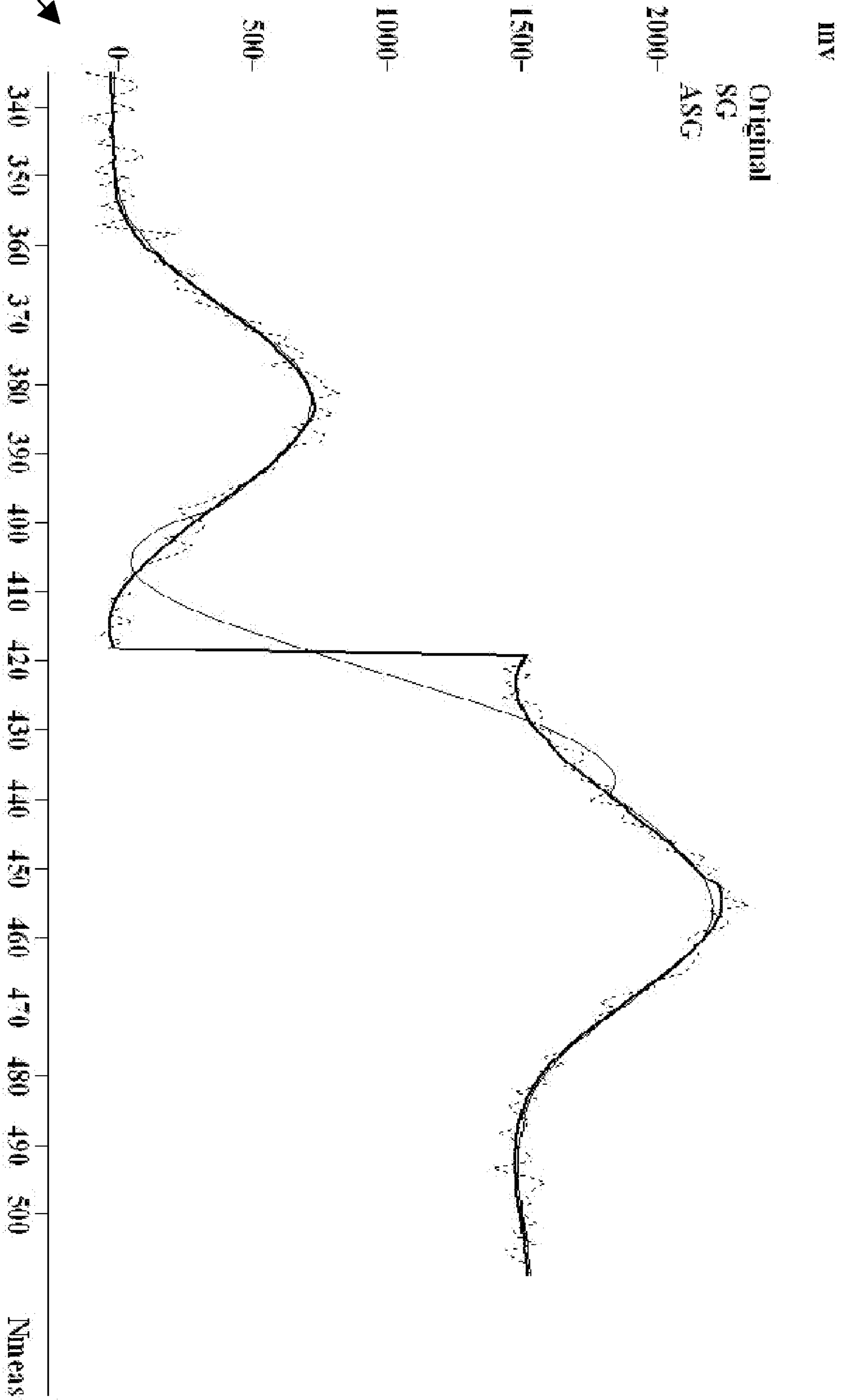
mv
0
500
1000
1500
2000
Original
SG
ASG
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
Nmeas

FIG. 7

800

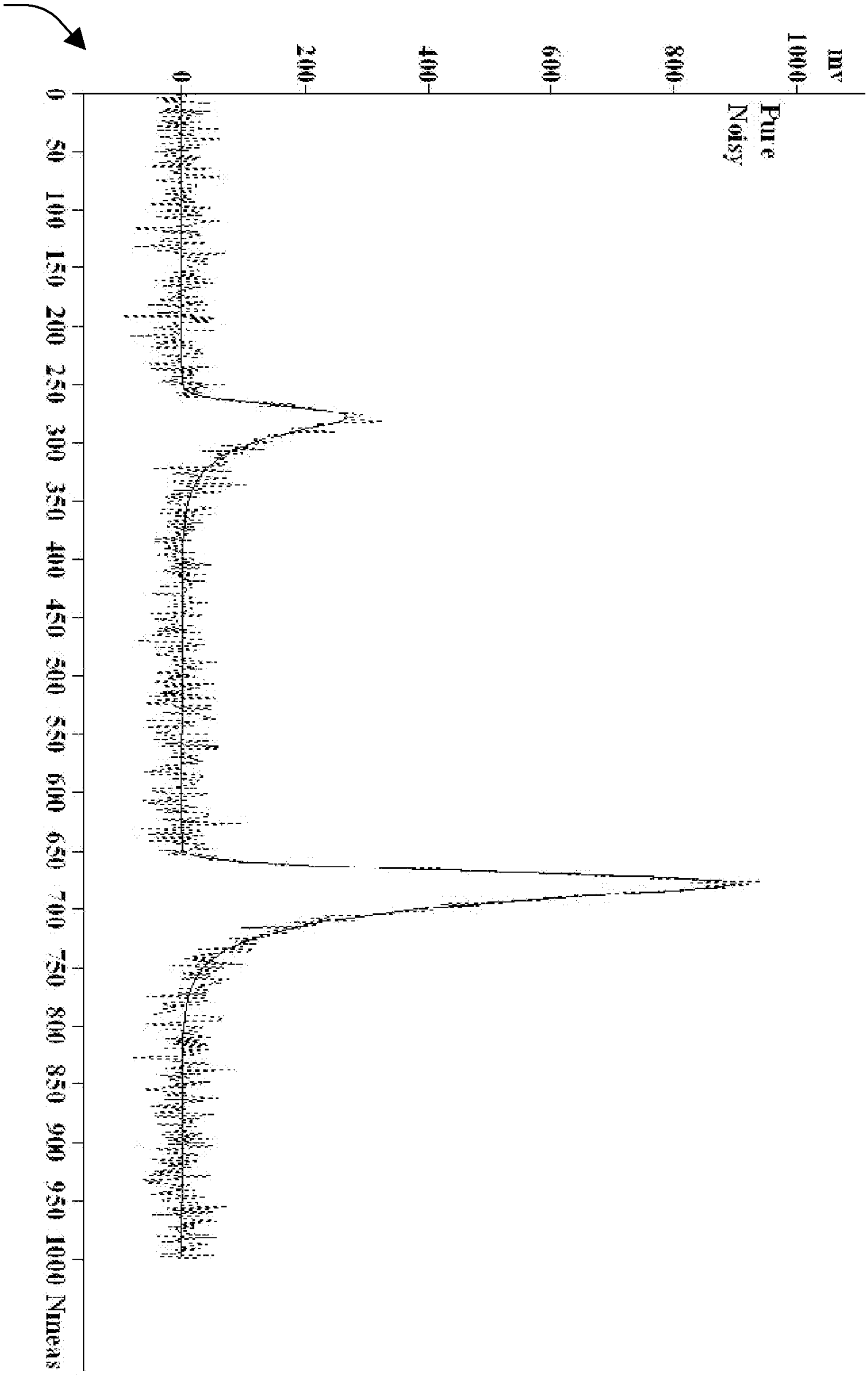
mv
1000
800
600
400
200
0
Pure
Noisy
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
Nmeas

FIG. 8

900

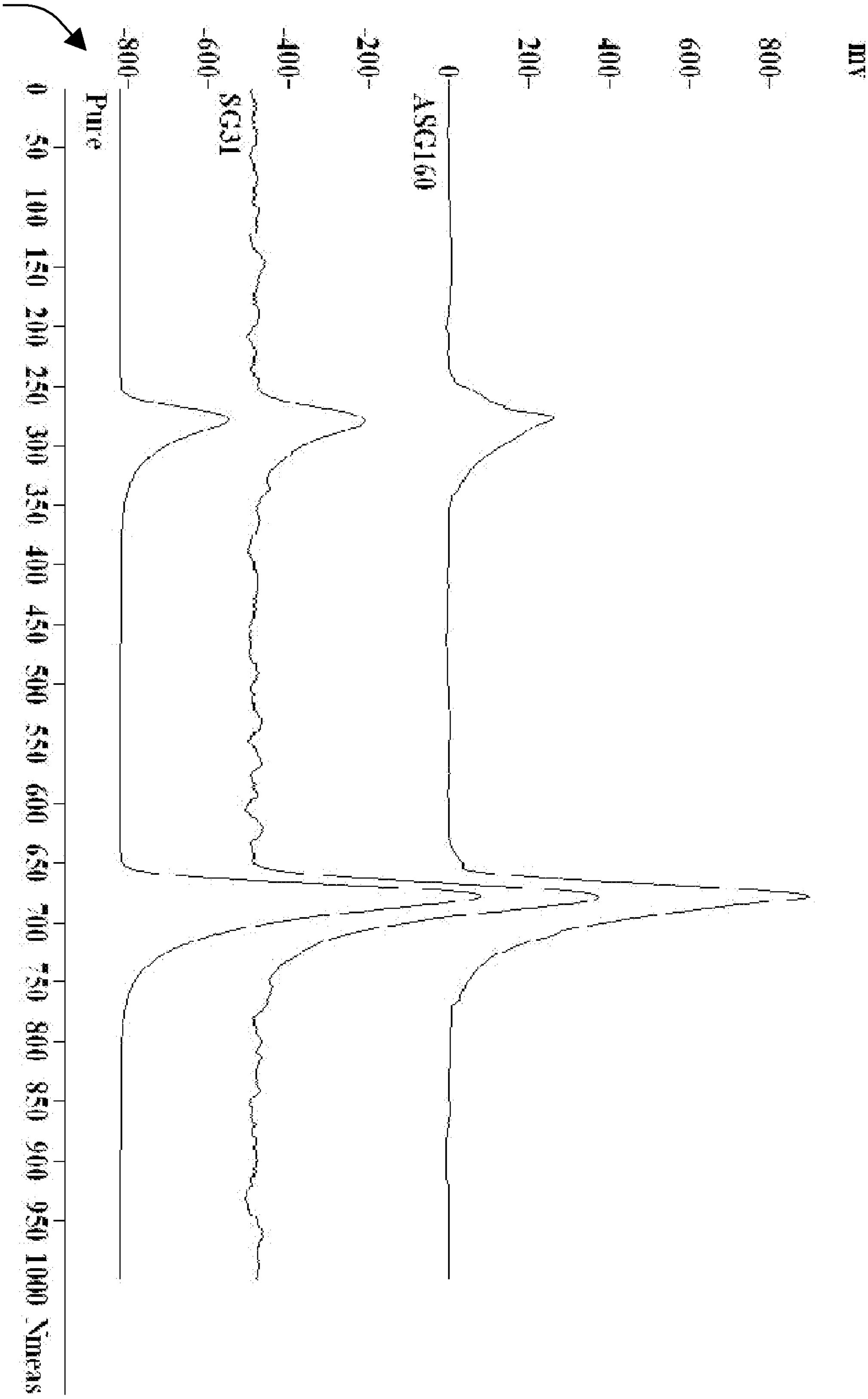
mv
800
600
400
200
0
-200
-400
-600
-800
ASG160
SG31
Pure
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000
Nmeas

FIG. 9

1000

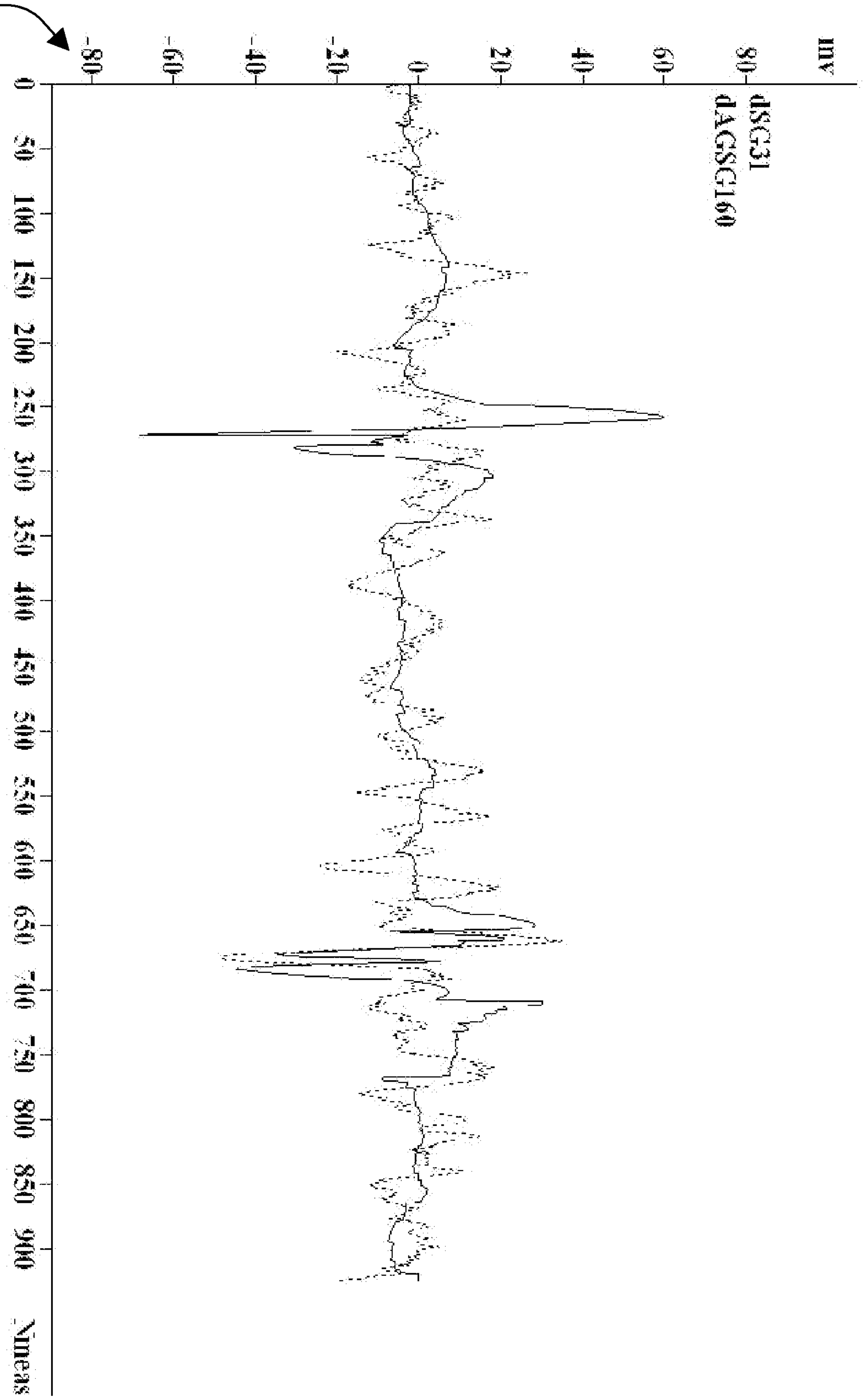
-80
-60
-40
-20
0
20
40
60
80
mv
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
Nmeas
dSG31
dAGSG160

FIG. 10

1100

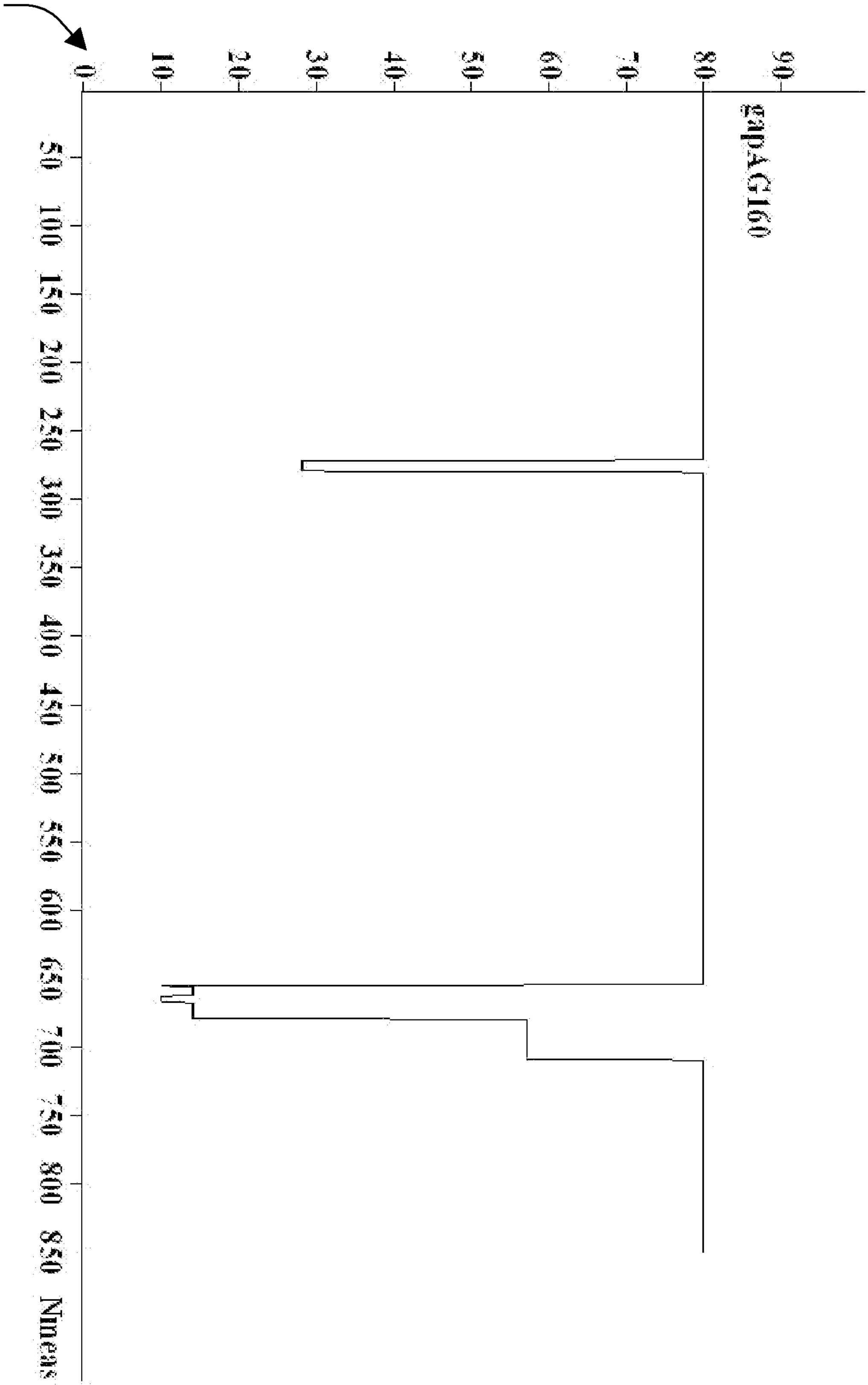
0
10
20
30
40
50
60
70
80
90
gapAG160
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
Nmeas

FIG. 11

1200

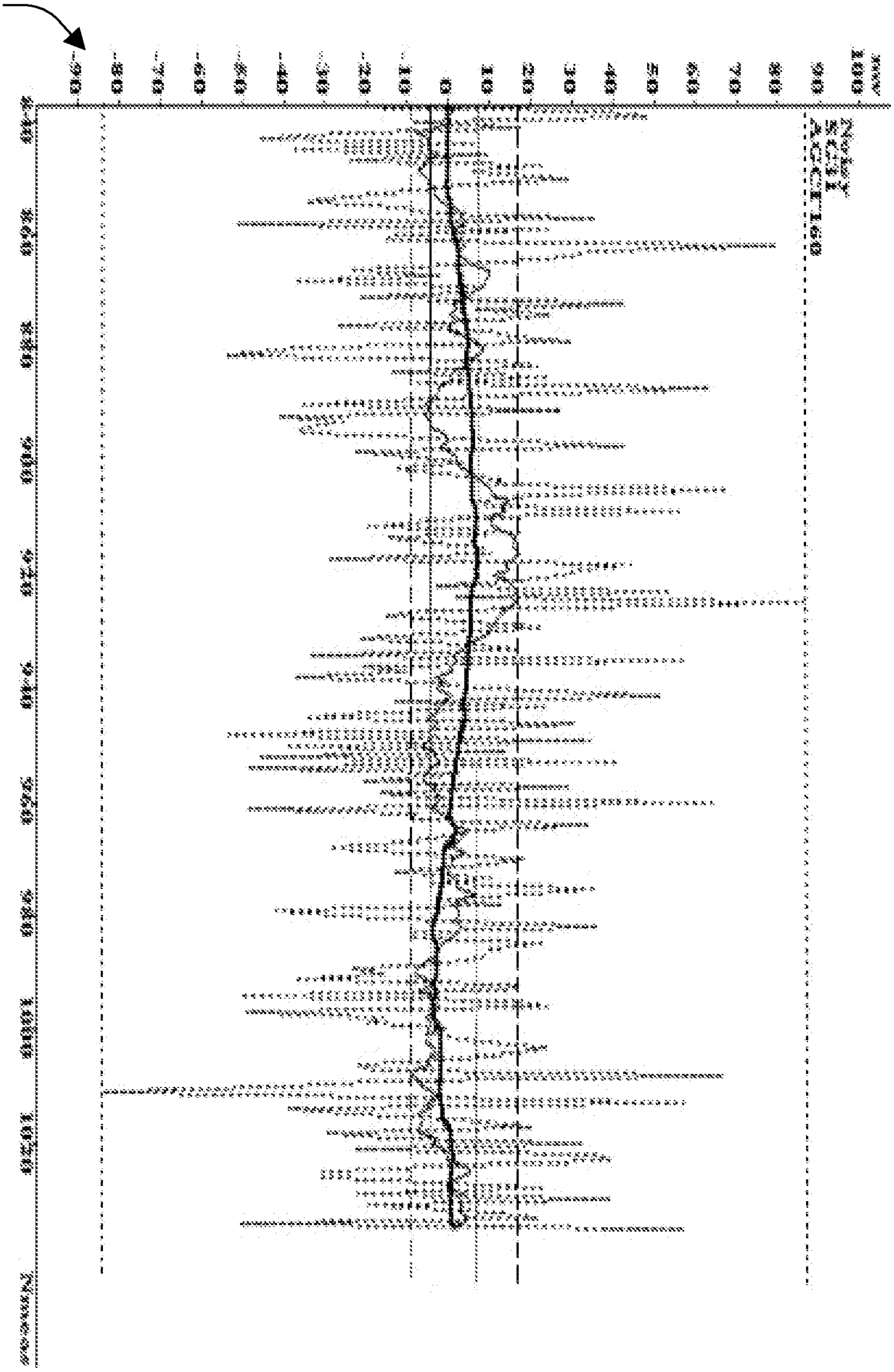

FIG. 12

1300

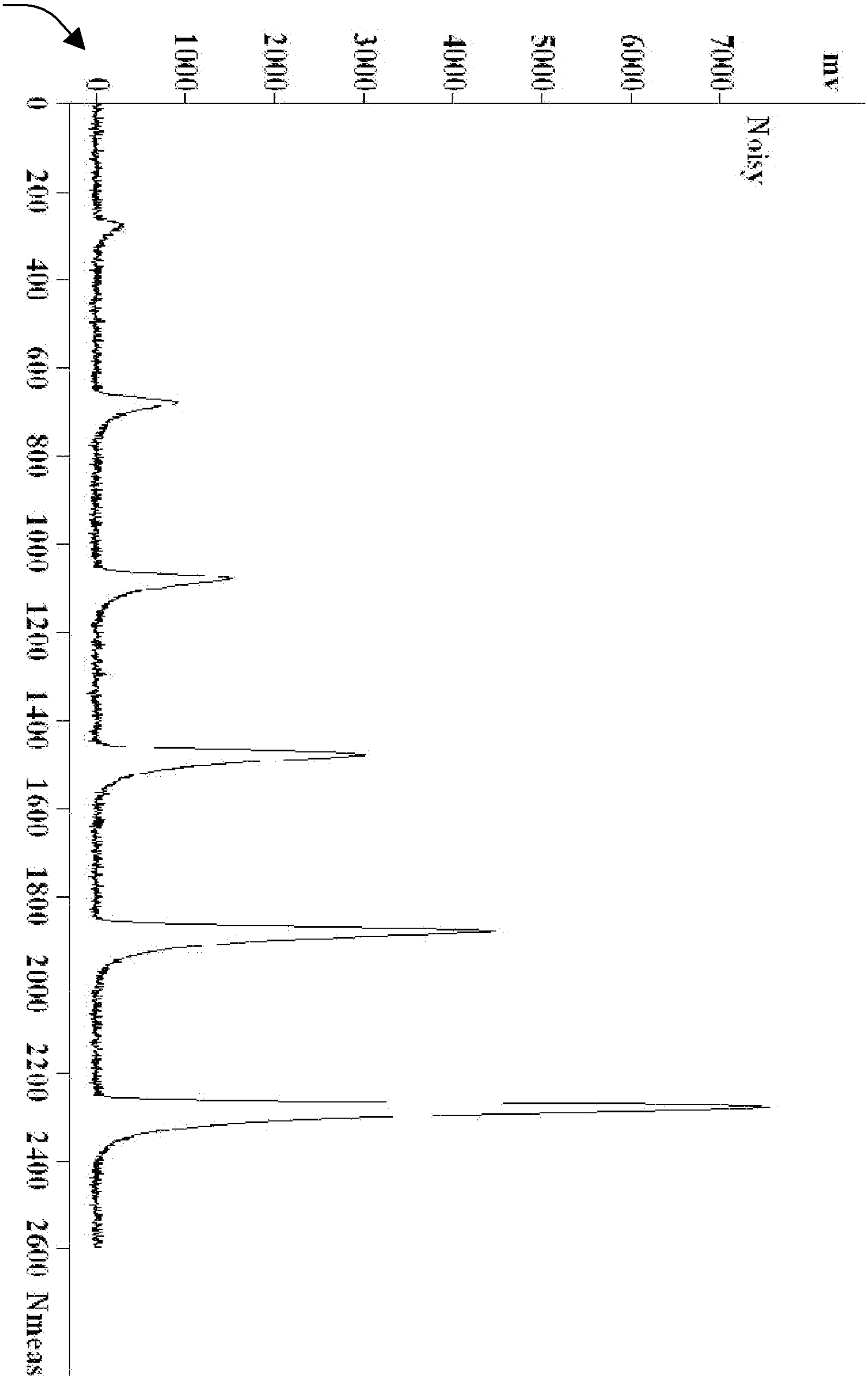
Noisy
mv
Nmeas
0
1000
2000
3000
4000
5000
6000
7000
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600

FIG. 13

1400

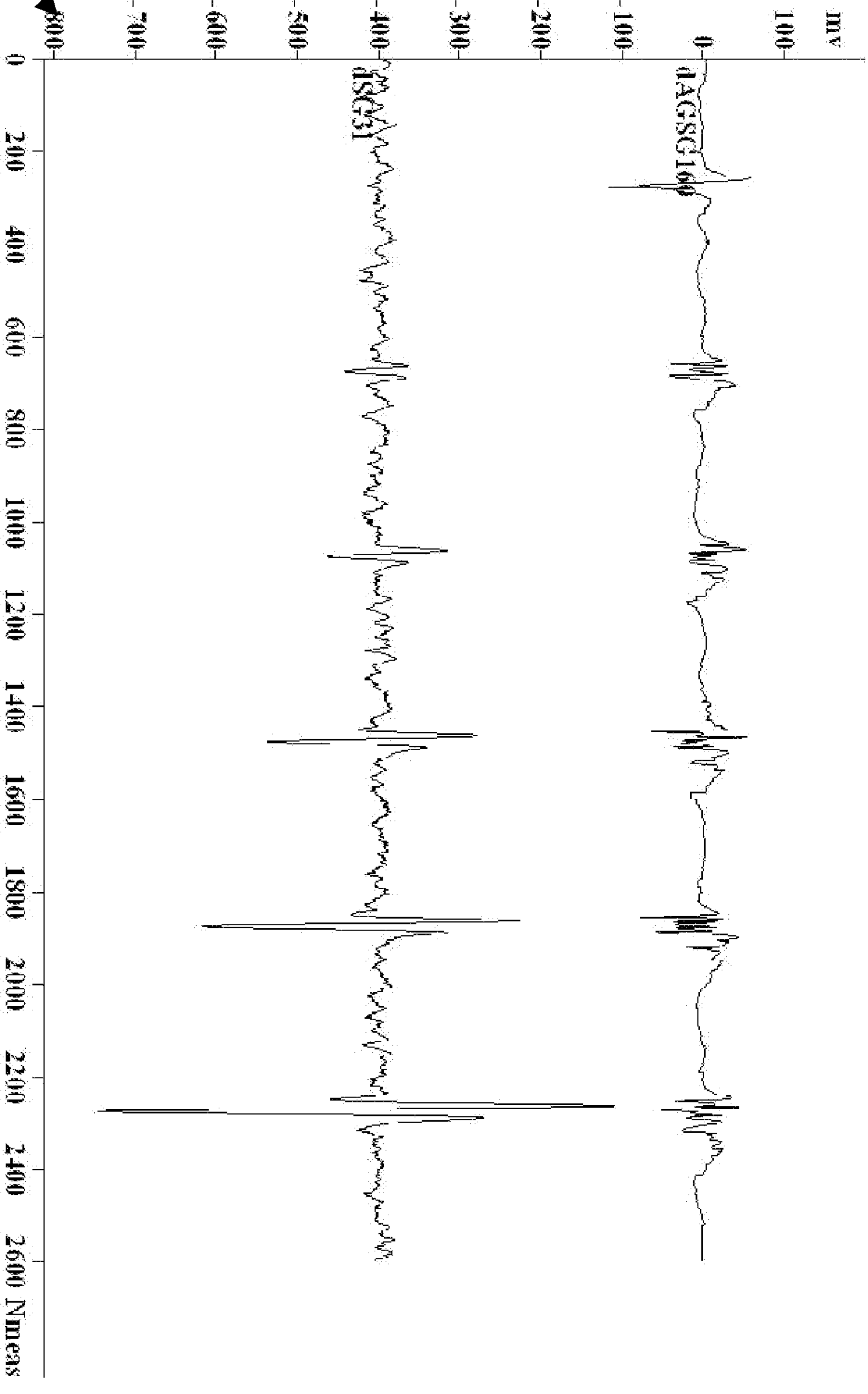
mV
dSG31
dAGSG160
Nmeas
-800
-700
-600
-500
-400
-300
-200
-100
0
100
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600

FIG. 14

1500

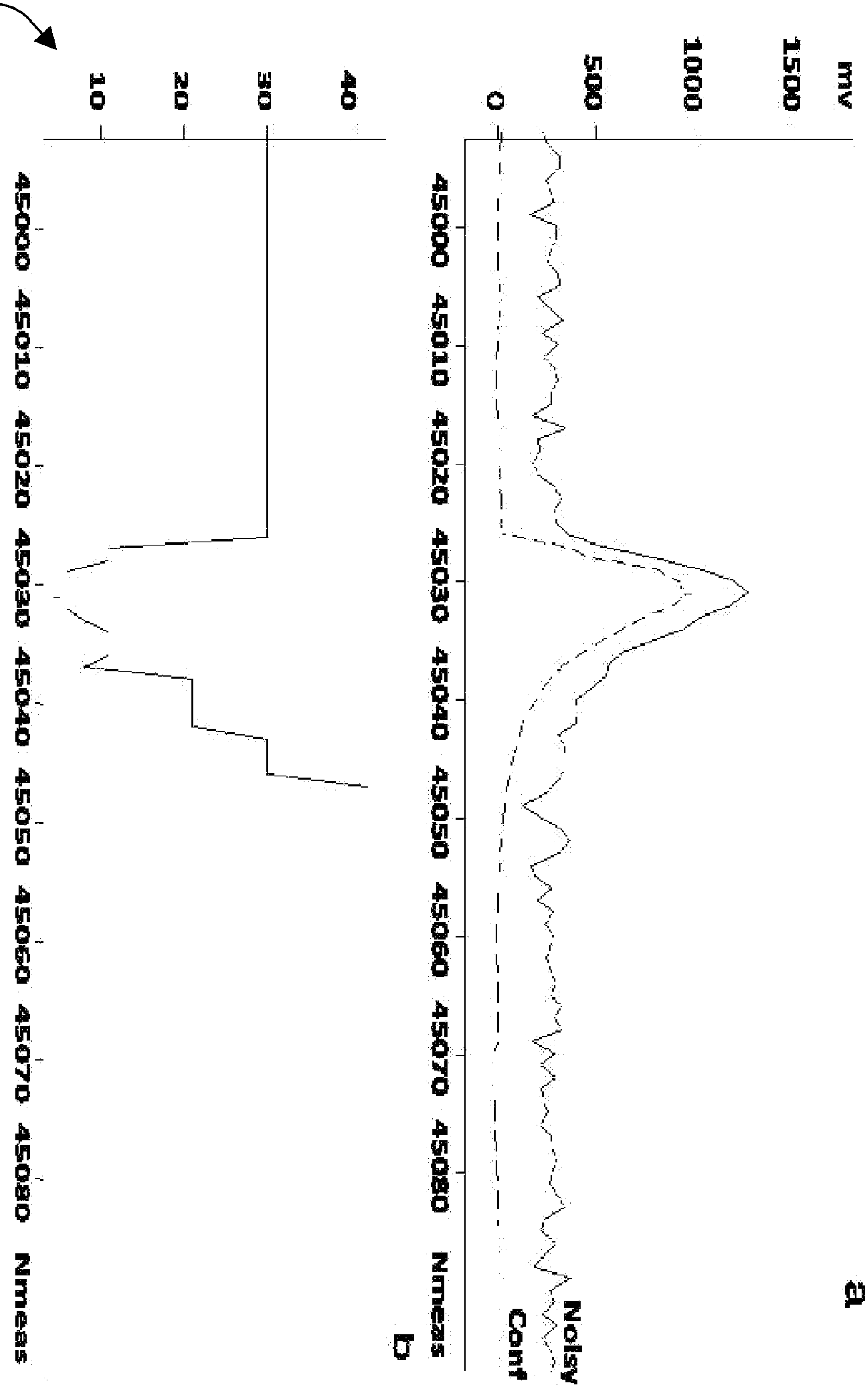
mv
1500
1000
500
0
45000 45010 45020 45030 45040 45050 45060 45070 45080 Nmeas
Noisy
Conf
a
b
10
20
30
40
45000 45010 45020 45030 45040 45050 45060 45070 45080 Nmeas

FIG. 15

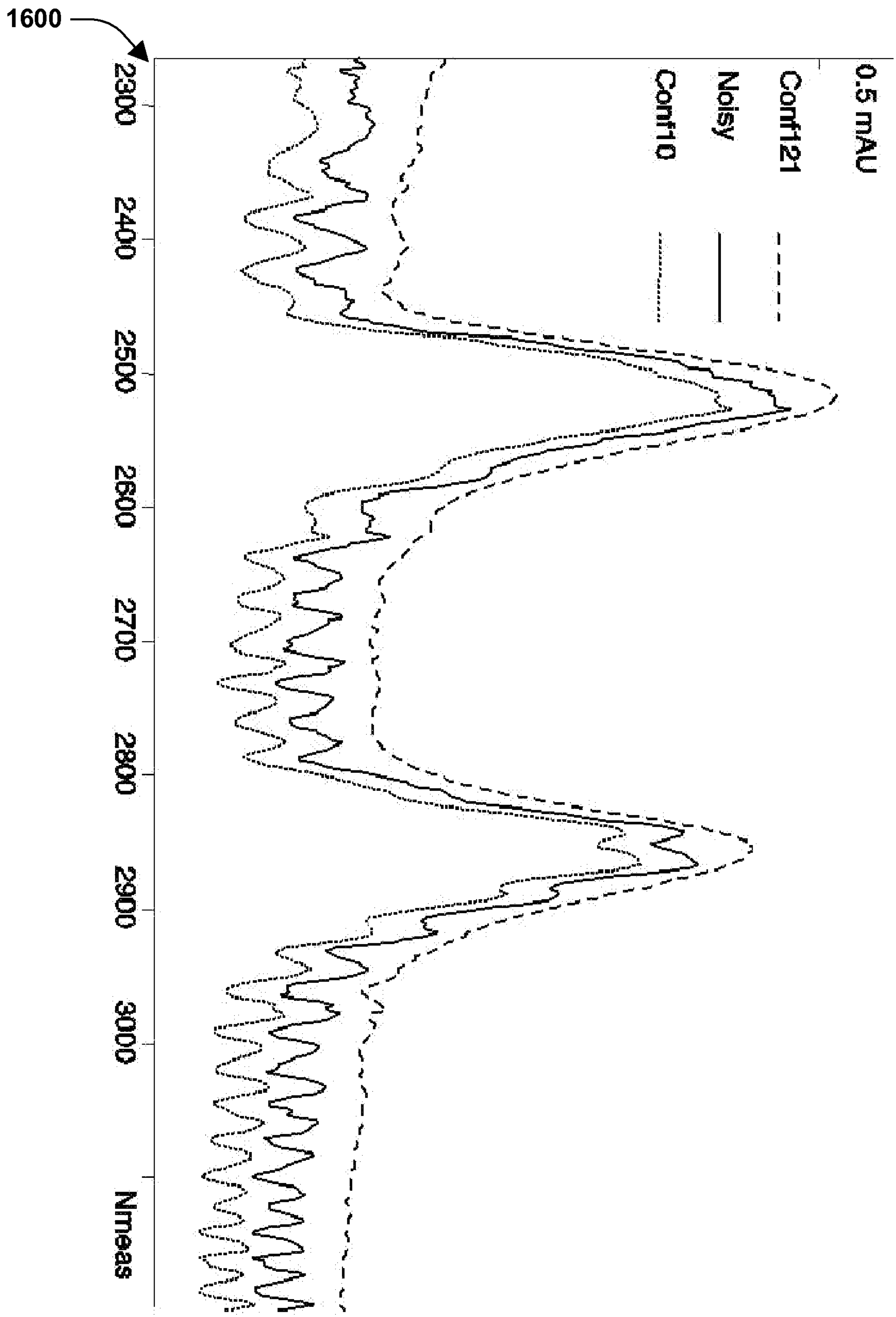
1600
0.5 mAU
Conf121
Noisy
Conf10
2300
2400
2500
2600
2700
2800
2900
3000
Nmeas

FIG. 16

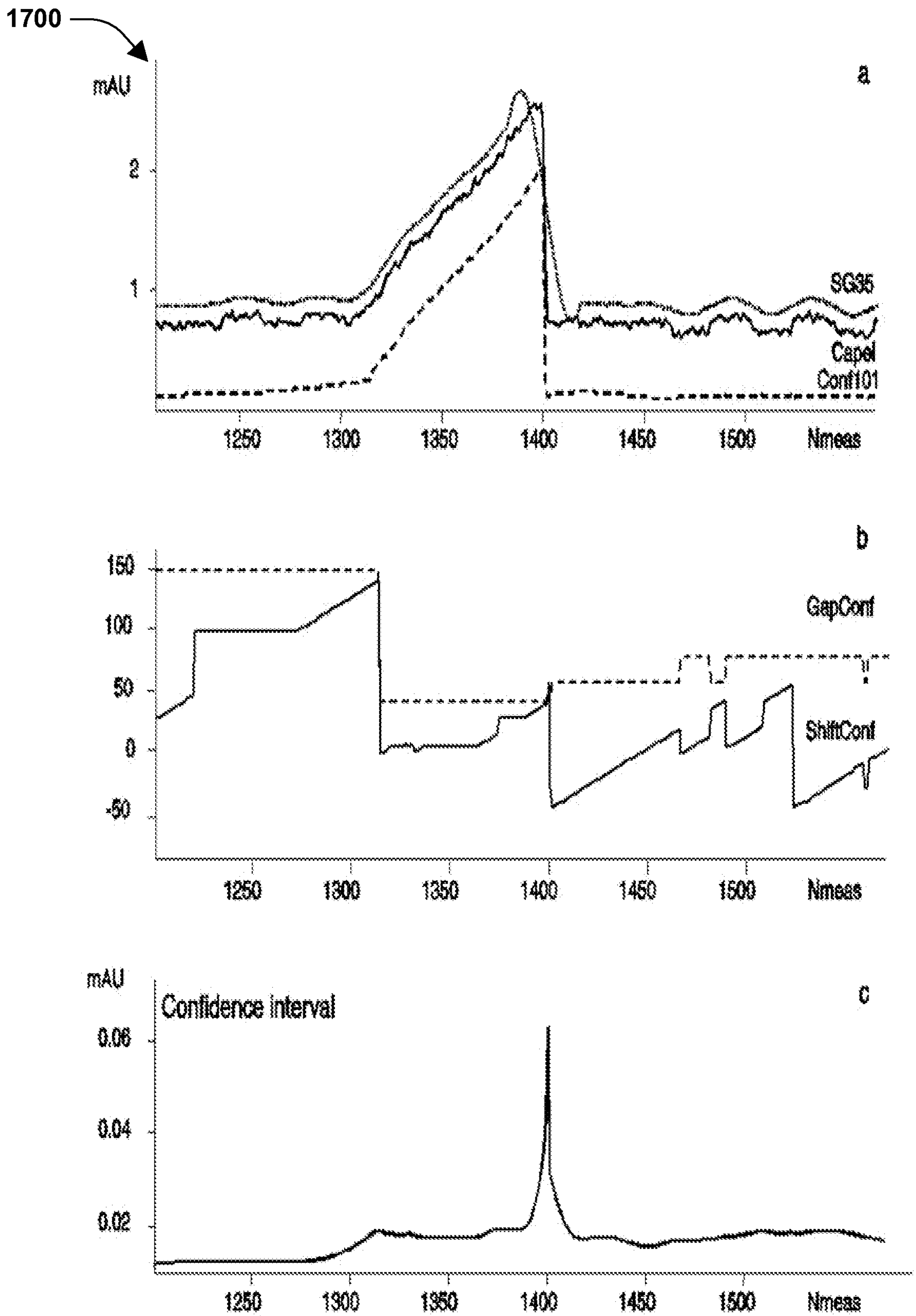
1700
a
mAU
2
1
SG35
Capel
Conf101
1250
1300
1350
1400
1450
1500
Nmeas
b
150
100
50
0
-50
GapConf
ShiftConf
1250
1300
1350
1400
1450
1500
Nmeas
c
mAU
Confidence interval
0.06
0.04
0.02
1250
1300
1350
1400
1450
1500
Nmeas

FIG. 17

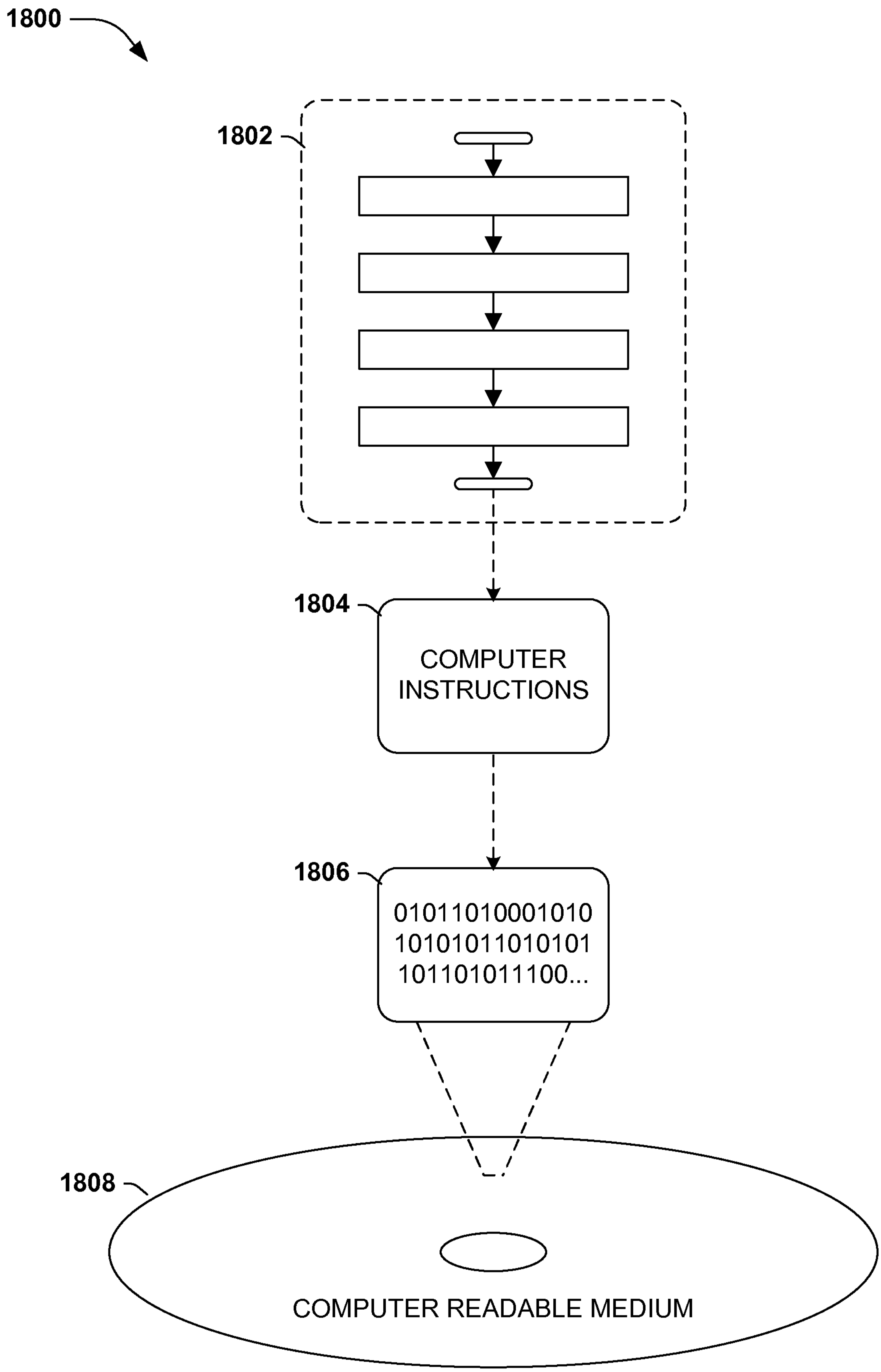
1800
1802
1804
COMPUTER INSTRUCTIONS
1806
01011010001010
10101011010101
101101011100...
1808
COMPUTER READABLE MEDIUM

FIG. 18

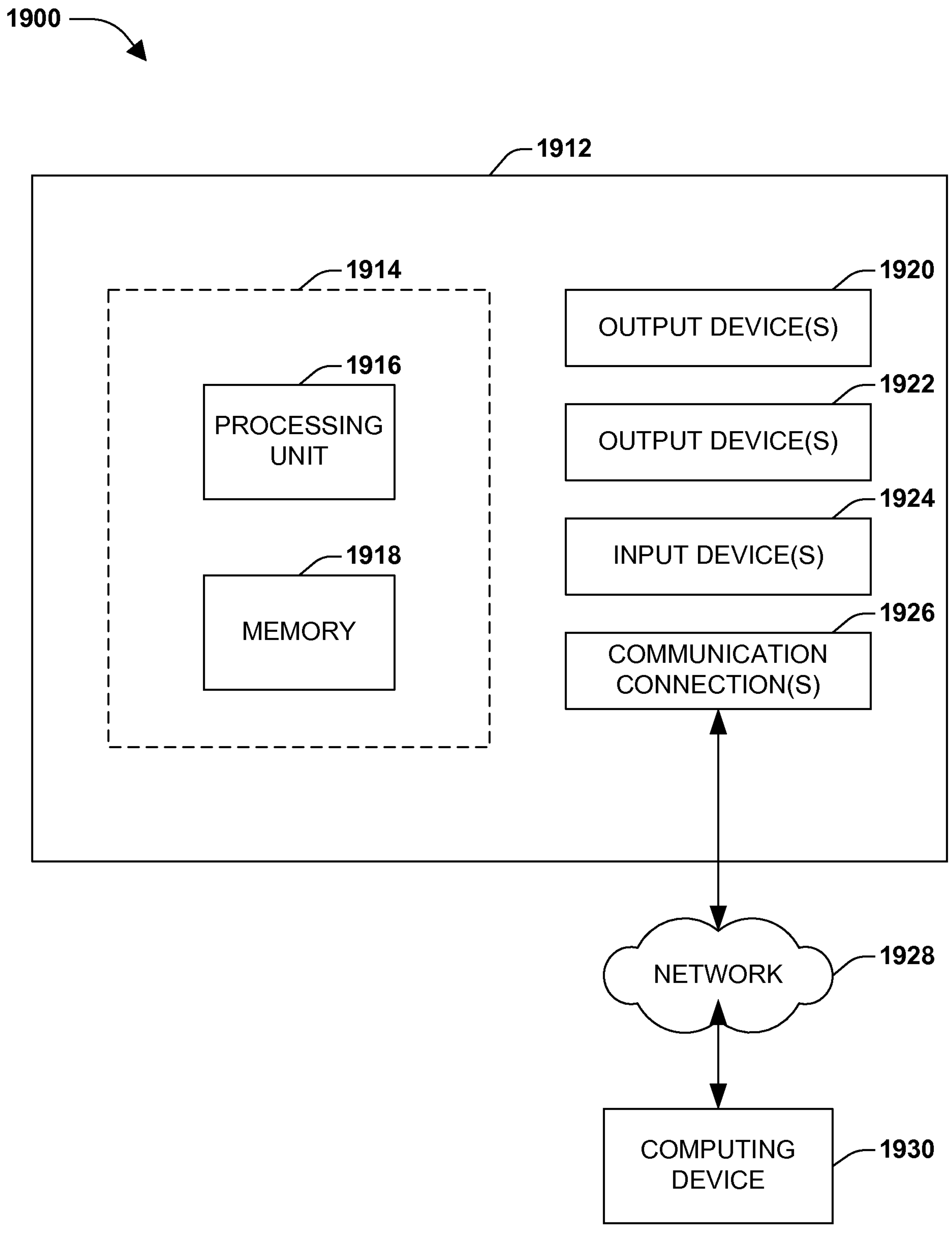
1900
1912
1914
1916
PROCESSING UNIT
1918
MEMORY
1920
OUTPUT DEVICE(S)
1922
OUTPUT DEVICE(S)
1924
INPUT DEVICE(S)
1926
COMMUNICATION CONNECTION(S)
NETWORK
1928
COMPUTING DEVICE
1930

FIG. 19