

B Mathematische Verfahren

Regressionsverfahren nach Levenberg-Marquardt

Die Methode von Levenberg-Marquardt [114, 121, 123] ist eine der meistgenutzten Algorithmen zur Anpassung von nichtlinearen Problemen. Diese Methode arbeitet mit einer Kombination aus der Methode des stärksten Abstiegs (steepest descent method) und der Linearisierungsmethode (Gauss-Methode), da keine der aufgeführten Methoden für sich selbst in jedem Fall zum Erfolg führt. Während die Methode des stärksten Abstiegs besser arbeitet, wenn sich der Betrachtungsraum weit entfernt vom Minimum befindet, liefert die Linearisierungsmethode bessere Ergebnisse in der Nähe des Minimums. Wichtig für den Umgang mit der Levenberg-Marquardt Verfahren ist die geeignete Wahl einer ausreichend kleinen Schrittweite und die geeignete Wahl an Startparametern. Die Theorie garantiert die Konvergenz hin zum Minimum (χ^2), was jedoch nicht in jedem Fall eine optimale Approximation an die vorgegebenen Daten beinhaltet. Für die Daten mit der abhängigen Variable y als Funktion der unabhängigen Variablen x und einem Vektor a mit der Anzahl an Parametern gilt.

$$\mathbf{y}_i = y(\mathbf{x}_i; \vec{a}) \quad (\text{B.1})$$

Das Anpassungsproblem besteht nun darin, diejenigen Parameterwerte zu finden, welche die vorliegenden Daten am besten mit dem zugrunde liegenden Modell beschreiben. Der Standardweg hierfür ist die Minimierung der Zielfunktion Chi-Square (χ^2) (Fehlerquadratsumme):

$$\chi^2(\vec{a}) = \sum_{i=1}^N \left(\frac{y_i - y(x_i; \vec{a})}{\sigma_i} \right)^2 \quad (\text{B.2})$$

N steht hierbei für die Anzahl der Datenpunkte und σ_i für die Standardabweichung. Diese Funktion bestimmt das Maß der Übereinstimmung zwischen den Datenpunkten und dem gewählten Anpassungsmodell. Je kleiner $\chi^2(\vec{a})$ wird, desto besser ist die Anpassung an das Modell.

- a) Methode des steilsten Abstiegs

Bei dieser Methode wird auf einer Ebene bzw. Hyperebene der Zielfunktion nach der Richtung des steilsten Abstiegs in Abhängigkeit von den Parametern und den Variablen gesucht. Dazu wird ein Versuchsplan aufgestellt, bei dem die Zielfunktion Chi mit Hilfe eines linearen Modells in den Parametern von a angepaßt wird. Die Richtung des steilsten Abstiegs ist durch die Werte

$$\chi^2 = a_0 + \sum_{i=1}^n a_j \frac{[b_j - \bar{b}_j]}{s_j} + e \frac{[b_j - \bar{b}_j]}{s_j} = \lambda(-a_j) \quad (\text{B.3})$$

gegeben. λ ist hierbei das Abstiegsparameter. Das Verfahren wird solange mit den neuen Parametern wiederholt, bis ein Minimum erreicht ist. Ist die Annäherung durch diese Methode nicht so genau, so erfolgt sie durch die

b) Linearisierungsmethode (Gauss-Verfahren) mittels Taylorreihen

Im Unterschied zu linearen Parametern lassen sich die nichtlinearen Parameter nicht mehr einfach als Produkte zwischen der unabhängigen Variablen und dem Parametervektor (vgl. Gleichung x) darstellen. Sind jedoch für den zu bestimmenden Parametervektor b genäherte Anfangswerte b_0 bekannt, kann die Funktion $y=f(x)$ so umgeschrieben werden, daß eine lineare Funktion in Abhängigkeit der Differenz des zu bestimmenden Parameters b und des Startwertes b_0 entsteht. Damit läßt sich dann das Verfahren auf die Lösung eines Kleinste-Quadrate-Problem zurückführen. Grundlage der Linearisierung ist die Entwicklung der Funktion in eine Taylor-Reihe. Dabei erfolgt unter der Annahme, daß die Differenz Δb genügend klein ist, der Abbruch nach dem Glied erster Ordnung:

$$y_i = f(x_i, b_1, b_2, \dots, b_m) + e_i \quad (\text{B.4})$$

$$y_i = f(x_i, b = b_0) + \left. \frac{\delta f_i}{\delta b_1} \right|_{b=b_0} \Delta b_1 + \left. \frac{\delta f_i}{\delta b_2} \right|_{b=b_0} \Delta b_2 + \dots + \left. \frac{\delta f_i}{\delta b_m} \right|_{b=b_0} \Delta b_m + e_i \quad (\text{B.5})$$

wobei

$$\left. \frac{\delta f_i}{\delta b_m} \right|_{b=b_0} \quad (\text{B.6})$$

die erste Ableitung nach dem Parameter b_j am Punkt i an der Stelle des Anfangswertes b_0 ist.

Die Gleichung ist damit linear in den Parametern und ähnlich der Gleichung für die lineare Regression (??). Die Matrix \mathbf{X} enthält jetzt jedoch die ersten

Ableitungen der Funktion.

$$\begin{pmatrix} \frac{\delta f_1}{\delta b_1} & \frac{\delta f_1}{\delta b_2} & \dots & \frac{\delta f_1}{\delta b_m} \\ \frac{\delta f_2}{\delta b_1} & \frac{\delta f_2}{\delta b_2} & \dots & \frac{\delta f_2}{\delta b_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta f_n}{\delta b_1} & \frac{\delta f_n}{\delta b_2} & \dots & \frac{\delta f_n}{\delta b_m} \end{pmatrix} = X = J \quad (\text{B.7})$$

Diese Matrix bezeichnet man in der nichtlinearen Regression als die Jacobi-Matrix, **J**. Das linearisierte Modell lautet damit:

$$\Delta \mathbf{y} = J \Delta \mathbf{b} + e \quad (\text{B.8})$$

Ist das Zielgebiet mit der gewählten Taylor-Reihe nicht ausreichend genug beschrieben, kann eine Erweiterung erfolgen, indem die Glieder der zweiten Ableitung noch mit einbezogen werden. Diese neue Matrix wird dann Hessen-Matrix, **H** genannt.

Bei dem Levenberg-Marquardt Verfahren werden folgende Schritte durchlaufen:

1. Berechnung von $\chi^2(a)$
2. Festlegung der Schrittweite: λ
3. Berechnung der Lösungen für δa .
4. Auswertung von $\chi^2(a + \delta a)$
5. Falls $\chi^2(a + \delta a) \geq \chi^2(a)$, dann erhöhe λ um einen Schritt und gehe zu Punkt 3 zurück.
6. Falls $\chi^2(a + \delta a) \leq \chi^2(a)$, dann verringere den Wert λ um einen Schritt, korrigiere die Parameter um $a = a + \delta a$ und gehe zu Punkt 3 zurück.
7. Wenn Falls $\chi^2(a + \delta a)$



Toleranz, dann beende die Iteration.

Diese Vorgehensweise kann an beiden nachfolgenden Diagrammen (Abb.B.1 und B.2) veranschaulicht werden. In einem Koordinatensystem ist der Lösungsraum in Abhängigkeit von den beiden Variablen x und y dargestellt. Der Lösungsraum ist begrenzt durch die Chi-Square-Funktion (χ_0^2), die hier als Linie dargestellt ist. Diese wird gebildet durch die Wahl des Startpunktes (S_0) der Regression. Die erste Iteration entspricht einem Linearisierungsprozeß und liefert das Ergebnis

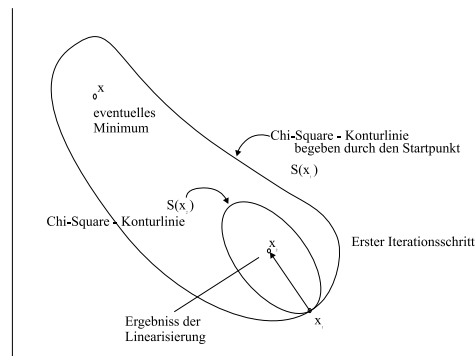


Abbildung B.1: Die erste Iteration in einem Linearisierungsprozeß.

S_1 mit der ellipsenförmigen Konturlinie der dazugehörigen neuen Chi-Square-Funktion (χ_1^2), welche an die Konturlinie der ersten Chi-Square-Funktion (χ_0^2) angrenzt (s. Abbildung B.1). Die Berechnung erfolgt nun von dem Punkt S_1 , dem Ergebnis der ersten Iteration, ausgehend. Weitere Iterationen liefern weitere Werte, bis letztlich das Minimum erreicht ist (s. Abbildung B.2).

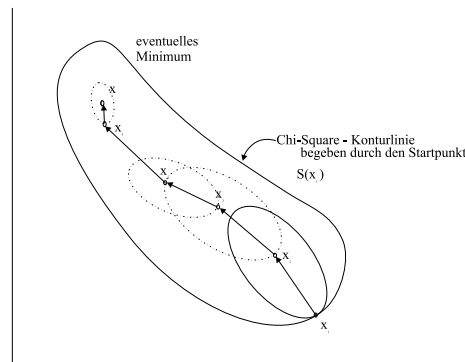


Abbildung B.2: Folgeiterationen in einem Linearisierungsprozeß.

Response Surface Methode

Weitaus komplizierter wird es, wenn die Meßgröße von zwei oder mehreren Einflußgrößen gleichzeitig abhängt. Im diesem Fall erfolgt die Regression im 3- bzw. n-dimensionalen Raum, unter Berücksichtigung einer Vielzahl von Möglichkeiten der Anpassung. Eine gängige Methode zur Beschreibung und Visualisierung dieser Abhängigkeiten ist die Antwortoberflächenmethode (response surface method). Zur Beschreibung des Zusammenhanges zwischen den Einflußgrößen und

der Meßgröße werden polynomische Funktionen unterschiedlicher Ordnung herangezogen [53, 122]. Das einfachste Modell zur Beschreibung der Abhängigkeit der Zielgröße von zwei Einflußgrößen ist ein lineares Modell erster Ordnung:

$$y_1 = b_0 + b_1x_1 + b_2x_2 \quad (\text{B.9})$$

Oft reicht dieses Modell zur Beschreibung jedoch nicht aus, da ebenfalls eine

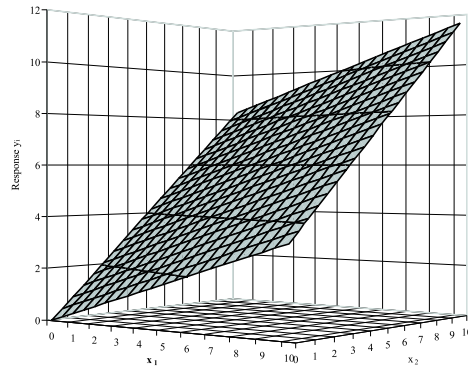


Abbildung B.3: Response Surface Modell Typ I auf Basis von zwei Einflußgrößen.

Wechselwirkung zwischen den Einflußgrößen denkbar ist. Das Modell wird durch folgende mathematische Beschreibung wiedergegeben:

$$y_1 = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 \quad (\text{B.10})$$

$$y_1 = b_0 + b_1x_1 + (b_2 + b_{12}x_1)x_2 \quad (\text{B.11})$$

Die Wechselwirkung zwischen den beiden Einflußgrößen wird hierbei durch den

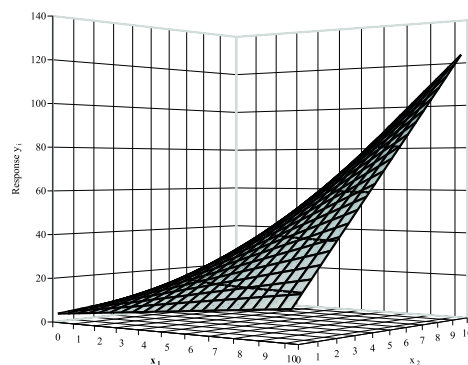


Abbildung B.4: Response Surface Modell Typ II auf Basis von zwei Einflußgrößen.

Summanden $b_{12}x_1x_2$ und das Ergebnis der ersten Ableitung beschrieben, da die

Steigung bezüglich x_2 von x_1 abhängt. Neben den Modellen erster Ordnung bieten sich oft polynomische Modelle zweiter Ordnung an, um die Variablenabhängigkeiten zu beschreiben. Neben den klassischen Modellen, die wie zuvor im Modell erster Ordnung auf die Wechselwirkung zwischen den einzelnen Einflußgrößen verzichten, gibt es alle denkbaren Möglichkeiten zur Beschreibung der einseitigen und beidseitigen Abhängigkeit der Variablen untereinander. Für den klassischen Fall einer solchen polynomische Abhängigkeit gilt:

$$y_1 = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 \quad (\text{B.12})$$

Die beidseitige Abhängigkeit zwischen x_1 und x_2 wird in Analogie zu B.10 wie

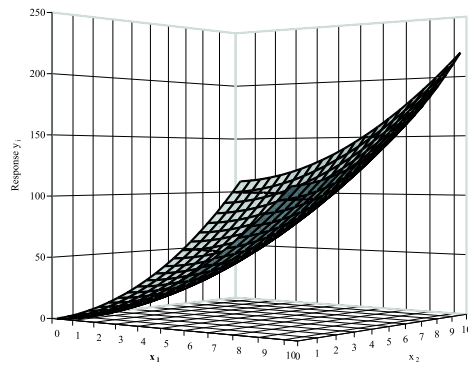


Abbildung B.5: Response Surface Modell Typ III auf Basis von zwei Einflußgrößen.

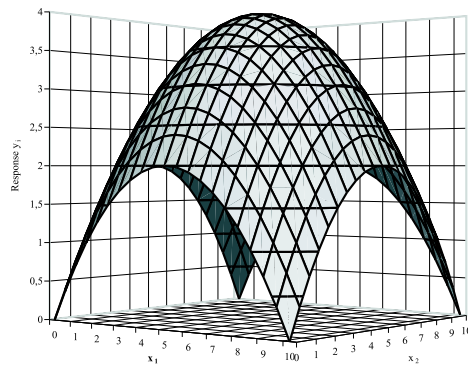


Abbildung B.6: Response Surface Modell Typ IV auf Basis von zwei Einflußgrößen.

folgt beschrieben:

$$y_1 = b_0 + (b_1 + b_{11}x_1)x_1 + (b_2 + b_{22}x_2)x_2 + b_{12}x_1x_2 \quad (\text{B.13})$$

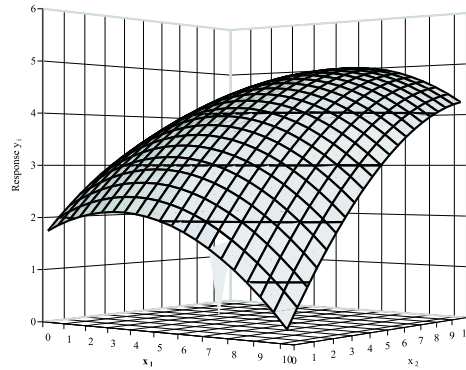


Abbildung B.7: Response Surface Modell Typ V auf Basis von zwei Einflußgrößen.

Auch denkbar ist weiterhin die einseitige Abhängigkeit zwischen den Variablen x_1 und x_2 . Da bei einigen Problemstellungen die Beschreibung mit Polynomen zweiter Ordnung nicht ausreicht, um die Abhängigkeiten exakt genug darzustellen, ist eine Erweiterung auf Polynome höherer Ordnung jederzeit möglich. Der Möglichkeit sind keine Grenzen gesetzt, jedoch sollte bei der Wahl des jeweiligen Modells der Realitätsbezug weitestgehend berücksichtigt werden. Die Bestimmung der einzelnen Regressionskoeffizienten erfolgt mittels des K-Matrix-Verfahrens (kleinste Quadrate Regression - ordinary least squares (OLS)). Für die aus n Wiederholungsmessungen stammenden Meßwerte und Einflußgrößen gilt:

$$\begin{aligned}
 y_{11} &= b_{10} + b_{11}x_{11} + b_{12}x_{12} + b_{13}x_{11}x_{12} + b_{14}x_{11}^2 + b_{15}x_{12}^2 \\
 y_{21} &= b_{20} + b_{21}x_{21} + b_{22}x_{22} + b_{23}x_{21}x_{22} + b_{24}x_{21}^2 + b_{25}x_{22}^2 \\
 &\vdots \\
 y_{n1} &= b_{n0} + b_{n1}x_{n1} + b_{n2}x_{n2} + b_{n3}x_{n1}x_{n2} + b_{n4}x_{n1}^2 + b_{n5}x_{n2}^2
 \end{aligned}
 \tag{B.14}$$

Diese allgemeine Darstellung läßt sich ebenfalls in Form von Matrizen darstellen:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{11}x_{12} & x_{11}^2 & x_{12}^2 \\ x_{21} & x_{22} & x_{21}x_{22} & x_{21}^2 & x_{22}^2 \\ \vdots & & & & \\ x_{n1} & x_{n2} & x_{n1}x_{n2} & x_{n1}^2 & x_{n2}^2 \end{pmatrix} \cdot \begin{pmatrix} b_{10} & b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{20} & b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ \vdots & & & & & \\ b_{n0} & b_{n1} & b_{n2} & b_{n3} & b_{n4} & b_{n5} \end{pmatrix}
 \tag{B.15}$$

Oder vereinfacht in der Matrixschreibweise:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}_y
 \tag{B.16}$$

Unter \mathbf{E}_y (Residuen) ist hierbei die Differenz zwischen den gemessenen und den mit dem Modell vorhergesagten y -Werten $\hat{\mathbf{Y}} - \mathbf{XB}$ zu verstehen. Da die Matrix

der unabhängigen Variablen \mathbf{X} nicht quadratisch ist, müssen die Regressionskoeffizienten \mathbf{B} über die generalisierte Inverse berechnet werden.

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X}^{-1}) \mathbf{X}^T \mathbf{Y} \quad (\text{B.17})$$

Der Vorhersageschritt lautet damit:

$$\mathbf{y} = \mathbf{xB} + \mathbf{E}_r \quad (\text{B.18})$$

Damit liegt die folgende Minimierungsaufgabe vor:

$$(\mathbf{y} - \mathbf{xB})^T (\mathbf{y} - \mathbf{xB}) = \text{Min!} \quad (\text{B.19})$$

Deren Lösung mit

$$\mathbf{x} = (\mathbf{B}^T \mathbf{B}^{-1}) \mathbf{B}^T \mathbf{y} \quad (\text{B.20})$$

beschrieben werden kann.

Kriging-Methoden

Eine andere Möglichkeit räumlich und/oder zeitlich korrelierte Daten zu analysieren, liefern geostatistische Methoden. Diese wurden für den Bergbau entwickelt, um die Lage von Lagerstätten beschreiben zu können. Erste Arbeiten gehen auf Krige [76], Wijs [109] und Matheron [120] zurück. Eine ausführliche Beschreibung der mathematischen Grundlagen ist in [60, 110, 125] zu finden. Die als Semivariogrammanalyse bekannte Korrelationsmethode ermöglicht die räumliche Darstellung der Abhängigkeiten von Probenwerten [115]. Der Vorteil dieser Methode gegenüber der Korrelationsanalyse besteht darin, daß bei unregelmäßig im Raum verteilten Probenpunkten eine Berechnung der Verbundenheit möglich ist. Basierend auf der Semivariogrammanalyse können ebenfalls zufällig verteilte Daten in ein reguläres Gitter umgewandelt werden. Dieses Verfahren wird „Kriging“ genannt. Die spezielle Korrelationsmethode nutzt die gewichtete räumliche Interpolation aus, um für jede Zelle einer regulären Matrix aus den Werten der Punkte der angrenzenden Zellen, welche innerhalb der Suchumgebung liegen, einen neuen Wert zu berechnen. Das genaue Verfahren ist bei Davis [60] aufgeführt. Hier sollen nur die wichtigsten Gleichungen besprochen werden. Für die Daten an den nicht untersuchten Stellen gilt beim einfachen Punkt-Kriging:

$$\hat{z}(x_o) = \sum_{i=1}^n [w_i z(x_i)] \quad (\text{B.21})$$

$\hat{z}(x_o)$ ist dabei der Schätzwert der Zufallsfunktion Z , an der nicht beobachteten Stelle x_o und w_i sind die Gewichte der Funktion $z(x_i)$. Die empirischen Gewichte

w_i werden mit Hilfe der Kriging-Gleichungen [110] ermittelt. Als Wichtungsfunktion dient das Semivariogramm der Gleichung B.22.

$$\sum_{i=1}^n w_i \gamma(x_i; x_j) + \mu = \gamma(x_i; x_0), i = 1, 2, \dots, n \quad (\text{B.22})$$

(x_i, x_j) ist dabei die Entfernung zwischen den Datenpunkten x_i und x_j und (x_i, x_0) der Abstand zwischen den gemessenen und den nicht gemessenen Datenpunkten, die es zu schätzen gilt. μ ist der Lagrange-Multiplikator. Um verzerrungsfreie Schätzungen zu erhalten, ist zu beachten, daß die Summe aller Gewichte gleich eins ist.

Für die Kriging-Varianz gilt:

$$\sigma_k^2 = \sum_{i=1}^n w_i \gamma(x_i; x_0) + \mu \quad (\text{B.23})$$

Die Güte der Anpassung läßt sich mittels der herkömmlichen Kreuzvalidierung ermitteln. Auch eine multivariate räumliche Anpassung mehrerer Variablen (Co-Kriging) ist möglich [125]. Abschließend sind zwei ausgewählte Beispiele für die räumliche Anpassung der Phasenlage und der Amplitude in Abhängigkeit der Konzentrationen von CO und CO₂ exemplarisch ausgewählt worden, um die prinzipiellen Vorgang des Krigings anhand von Abbildungen beschreiben zu können. Ausgehend von der räumlichen Verteilung der Datenpunkte in Anhängigkeit der beiden Konzentrationen von CO und CO₂ (s. Abb. B.8) sollen nun für eine reguläre Matrix die einzelnen Zellen aus den Werten der Punkte in den angrenzenden Zellen berechnet werden. Die Methode des gewichteten Durchschnitt beinhaltet

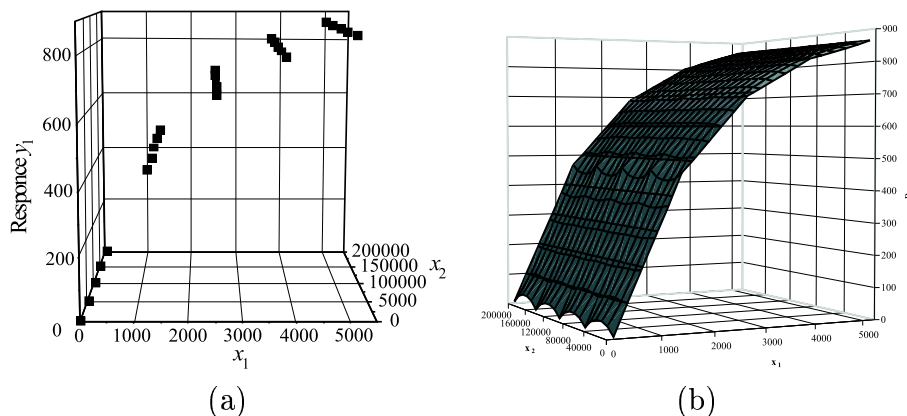


Abbildung B.8: 3D-Plot (a) und Surfaceplot (b) für die Verteilung der Datenpunkte bezüglich der Amplitude

tet einen einfachen gewichteten Mittelwert der Punkte, mit einem Gewicht $1/r$, wobei r der Abstand der Punkte von der Zelle ist. Liegt kein Punkt innerhalb des

gewählten Suchbereichs, wird der Suchradius solange erweitert, bis zumindest ein Punkt angetroffen wird. Die Erweiterung des Suchbereiches hat zur Folge, daß jeder Punkt enger mit den Nachbarpunkten in Beziehung steht. Dadurch ergibt sich eine glattere Oberfläche, in der keine Details verloren gehen. Dieses Verfahren wurde bei allen dreidimensionalen Oberflächendarstellungen angewendet. Auch bei der Bestimmung der Zusammensetzung des Detektors wurde ähnlich verfahren. Hier erfolgte jedoch die anschließende Darstellung in Form von Höhenlinien. Dazu später an geeigneter Stelle mehr. Anzumerken sind jedoch die gerade im Randbereich auftretenden Fehler (s. Abb. B.8(b)) bei der Durchführung dieses Verfahrens. Diese Fehler lassen sich durch eine höhere Anzahl an Datenpunkten verringern.

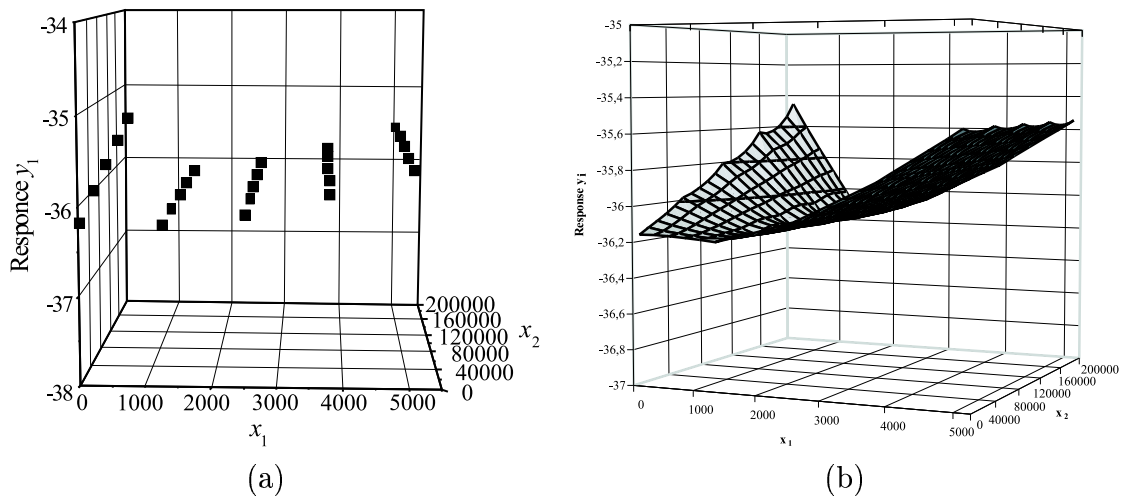


Abbildung B.9: 3D-Plot (a) und Surfaceplot (b) für die Verteilung der Datenpunkte bezüglich der Phase