

Genetische Algorithmen und Simulated Annealing: Nichtlineare Optimierung am Beispiel der Widerstandsgeoelektrik

Christoph Schwarzbach und Ralph-Uwe Börner¹

TU Bergakademie Freiberg

1 Einleitung

Neben den bekannten klassischen Verfahren, die bei der Inversion geophysikalischer Daten zur Minimierung eines Fehlerfunktional eingesetzt werden, existiert eine Reihe weiterer Ansätze, die von grundsätzlich anderen Strategien zur Inversion geophysikalischer Daten ausgehen. Im Folgenden sollen zwei dieser Verfahren, die genetischen Algorithmen sowie das Verfahren des Simulated Annealing, vorgestellt werden. An einem einfachen Datensatz einer widerstandsgeoelektrischen Sondierungskurve soll die Leistungsfähigkeit der Verfahren untersucht werden. Wir beschränken uns auf den Fall einer eindimensionalen Modellsituation, da bereits hier wesentliche Aussagen zum Verhalten der Algorithmen abgeleitet werden können.

Bei der Inversion, der Anpassung einer Modellantwort an einen Datensatz, wird ein Fehlerfunktional als Abstandsmaß zwischen Modellantwort und Daten bezüglich der Modellparameter minimiert. Das Fehlerfunktional hängt im Allgemeinen nichtlinear von den Parametern ab, die Abbildung vom Modell- in den Datenraum im Sinne eines Vorwärtsoperators existiert mit Ausnahme einfachster Modellklassen (z.B. des homogenen Halbraums, des vertikalen Kontakts, der vertikalen Platte) häufig nur als numerische Routine.

Die klassischen Verfahren zur Minimierung erfordern gute Startmodelle, da bei schlechten Startvorgaben Konvergenz gegen lokale Minima eintreten kann. Weiterhin problematisch ist die Berechnung der Sensitivitätsmatrix als Jacobi-Matrix des Vorwärtsoperators. Häufig, wie im Beispiel der Widerstandsgeoelektrik, liegt keine geschlossene Form der Fréchet-Ableitung für allgemeine Fälle vor. Darüberhinaus ist ihre Berechnung insbesondere für große Parameterräume aufwendig. Minimierungsverfahren, die keine Ableitungen benötigen, benutzen unterschiedliche Suchstrategien, um ein globales Minimum zu finden. Neben einer systematischen und damit zeitaufwendigen Suche bieten sich stochastische oder Monte-Carlo-Suchstrategien an.

Derartige nichtlineare Optimierungsalgorithmen mit stochastischer Suchstrategie besitzen den Vorzug, ein globales Minimum trotz Vorhandensein möglicher Nebenminima zu finden. Bekannteste Vertreter sind genetische Algorithmen sowie Simulated-Annealing-Verfahren. Beide Varianten nutzen Analogien zu natürlichen Prozessen, die optimale Zustände innerhalb von geschlossenen Systemen anstreben. Bei genetischen Algorithmen ist dies die Anpassung von Lebewesen an ihre Umwelt, Evolution und das Überleben der bestangepassten Individuen. Simulated Annealing basiert auf thermodynamischen Prozessen beim langsamen Abkühlen einer Schmelze, die zum Zustand minimaler Energie im Kristallgitter führen.

2 Genetische Algorithmen

Genetische Algorithmen (GA) als globale Suchverfahren setzen Strategien aus der Evolutionstheorie ein, um ein optimales Modell zu erhalten. Das Gütemaß (*Fitness*) als Funktion des Fehlerfunktional ist die einzige Information, die der GA zur Beurteilung eines Parametersatzes benötigt. Der GA

¹cschwarz@student.tu-freiberg.de, rub@geophysik.tu-freiberg.de

arbeitet mit einer Anzahl zu *Strings* codierten Parametersätzen, die eine *Population* im Modellraum bilden. Die genetischen Operatoren *Selektion*, *Kreuzung* und *Mutation* erzeugen aus den Strings einer zufallsgenerierten Startpopulation die jeweils nächste Generation, bis fast alle Strings um einen Punkt konzentriert sind (Konvergenz). Die genetischen Operatoren basieren auf stochastischen Prozessen und führen auf Grund ihrer Wirkungsmechanismen im Parameterraum zu unterschiedlichen Konsequenzen (vgl. Tab. 1 und Abb. 1).

Selektion	Kreuzung	Mutation
Für die Reproduktion werden bevorzugt diejenigen Strings ausgewählt, die eine hohe Fitness besitzen. Die Suche konzentriert sich auf erfolgversprechende Bereiche des Modellraums.	Zwei Strings werden zu zwei neuen Strings kombiniert, die Erbinformationen beider Elternteile enthalten. Kreuzung ist gemeinsam mit der Selektion für die globale Exploration des Modellraums wichtig.	Zufällige Änderung an einem String mit geringer Wahrscheinlichkeit. Mutation sichert die Diversität innerhalb der Population und bildet zusammen mit der Selektion die lokale Komponente bei der Suche.

Tabelle 1: Erläuterung der genetischen Operatoren Selektion, Kreuzung und Mutation.

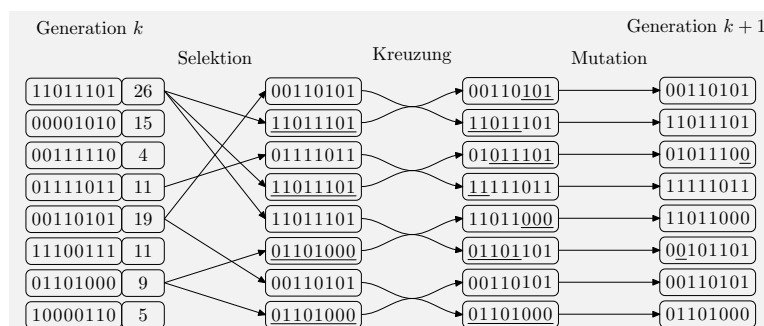


Abbildung 1: Selektion, Kreuzung und Mutation an bitkodierte Strings.

Die Strings einer Population können als *Muster* für alle ihnen ähnliche Strings angesehen werden. Damit bewertet der GA nicht den einzelnen String, sondern eine ganze Klasse von Mustern. Selektion und Kreuzung rekombinieren erfolgreiche Muster, um bessere Strings zu erzeugen (Holland, 1975). Damit sind GA die effektivste Methode, große Modellräume zu durchsuchen. Sie garantieren jedoch im Vergleich zu lokalen Suchverfahren keine Konvergenz mit vorgegebener Genauigkeit. Sinnvoll scheinen deshalb Hybridverfahren, bei denen der GA zum Finden geeigneter Startmodelle für lokale Suchverfahren eingesetzt wird.

3 Simulated Annealing

Der Begriff *annealing* beschreibt den technischen Prozess, bei dem ein Material für eine festgelegte Zeitdauer einer erhöhten Temperatur ausgesetzt wird, um die Bildung metastabiler Zustände wie Gitterunregelmäßigkeiten, -fehlstellen o. ä. zu vermeiden. Eine niedrige Abkühlungsrate garantiert schließlich, dass ein Zustand minimaler Energie erreicht werden kann.

Grundlage der Methode des Simulated Annealing ist die Analogie zur Thermodynamik, insbesondere zum Verhalten von kristallisierenden Flüssigkeiten oder erstarrenden Schmelzen. Die Energie E eines Systems im thermischen Gleichgewicht gehorcht einer Boltzmann-Wahrscheinlichkeitsverteilung $p(E) \propto \exp(-E/kT)$ mit T als Temperatur sowie der Boltzmann-Konstanten k . Danach kann ein System selbst bei niedriger Temperatur T noch mit einer geringen Wahrscheinlichkeit einen hohen

Energiezustand einnehmen. Entsprechend ist die Chance, dass das System ein lokales Energieminimum wieder verlässt. Als geeignetes Analogon zum Energiebegriff in der Thermodynamik erscheint zunächst das Fehlerfunktional zur Beschreibung der Anpassungsgüte zwischen Modellantwort und Daten. Der Abstand einer Modellparameterverteilung vom optimalen (gesuchten) Parametersatz entspricht der temperaturabhängigen Energie von Teilchen in einem System. Zusätzliche Forderungen z.B. nach Glattheit des Modells oder Berücksichtigung von *a-priori*-Informationen können ohne weiteres implementiert werden. Damit wird die Güte der Anpassung durch eine sehr allgemeine *Straffunktion* beschrieben.

Die Energiedifferenz zweier Modellantworten $\Delta E = E(m^{l+1}) - E(m^l)$ für zwei zufällig gezogene Modellparametersätze m^{l+1} und m^l wird bei konstanter Temperatur zur Berechnung der Wahrscheinlichkeit $p = \exp(-\Delta E/T)$ benutzt. Ist die Energiedifferenz negativ, d.h. ist die Anpassung besser, wird p auf 1 gesetzt und der aktuelle Parametersatz akzeptiert. Andernfalls wird der Parametersatz nur mit der Wahrscheinlichkeit p akzeptiert. Mit der Vorschrift zur Berechnung des n -ten Temperaturschrittes $T_n = T_0 \exp(-C \cdot n^{1/N})$ kann die Geschwindigkeit der Abkühlung über die Anfangstemperatur T_0 , die Abkühlungsrate C sowie den Exponenten $1/N$ mit N als Anzahl der Modellparameter gesteuert werden.

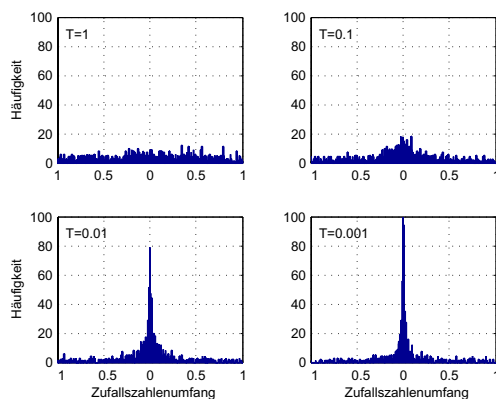


Abbildung 2: Histogramme Cauchy-verteilter Zufallszahlenfolgen für verschiedene Temperaturen T .

Der Very Fast Simulated Annealing Algorithmus (VFSA) nutzt zur Erzeugung der Parameteränderungen Cauchy-verteilte Zufallszahlen, welche über eine einfache Transformationsvorschrift aus gleichverteilten Zufallszahlen berechnet werden können. Die Form der Verteilungskurve ist temperaturabhängig (vgl. Abb. 2), wobei an beiden Enden der Kurven im Vergleich zu einer Normalverteilung größere Häufigkeiten zu beobachten sind. Das ermöglicht auch bei niedrigen Temperaturen noch große Parametersprünge, die eine kurzfristige Energieerhöhung des Systems und damit ein Verlassen lokaler Minima herbeiführen können.

4 Vergleich der Verfahren

Für ein 3-Schichtmodell wurde ein synthetischer Datensatz einer Schlumberger-Sondierung mit 1% normalverteiltem Rauschen erstellt (Abb. 3, links). Das zu minimierende Fehlerfunktional entspricht einer normierten χ^2 -Verteilung

$$E(m_1, \dots, m_p) = \frac{1}{\delta} \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n \frac{f_i(m_1, \dots, m_p)}{d_i} - 1 \right)^2}$$

mit $d_i, i = 1 \dots n$ als Datenvektor und $f_i, i = 1 \dots n$ als Vorwärtsoperator, angewendet auf den Parametersatz $m_j, j = 1 \dots p$.

Im Rahmen der Datengenauigkeit gibt es unendlich viele Modelle, die die Daten anpassen (Abb. 3, rechts: alle Punkte unterhalb des Fehlerniveaus von $E = 0.838$). In diesem Fall sind das Modelle mit $h_2/\rho_2 \approx 0.1$ S und $\rho_2 \approx 40 \Omega \cdot \text{m}$. Das globale Minimum für den gestörten Datensatz stimmt i. d. R. nicht mit dem Minimum des ungestörten Datensatzes überein.

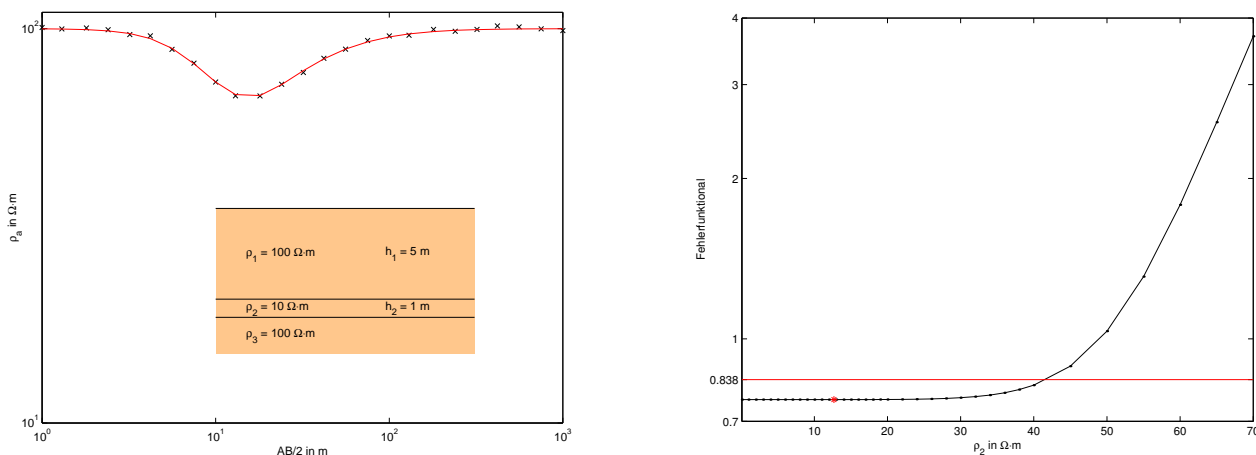


Abbildung 3: Links: Sondierungskurven über einem 1D-Modell für fehlerfreie (durchgezogene Linie) und verrauschte Daten (\times). Fehler normalverteilt mit $\delta = 1\%$. Rechts: Fehlerfunktional E minimiert bezüglich ρ_1, ρ_3, h_1, h_2 mit dem Parameter ρ_2 .

Zum Test der Algorithmen wurden folgende Programme weiterentwickelt bzw. neu geschrieben:

- MATLAB-Toolbox GAOT von Houck et al. (1996)
- MATLAB-Implementierung eines VFSA nach Ingber (1989)

Damit bestand die Möglichkeit, in einer konsistenten Testumgebung beide Verfahren hinsichtlich ihres Konvergenzverhaltens und der erreichten Anpassungsgüte zu untersuchen. Die invertierten Modelle sind für beide Verfahren in Tab. 2 aufgeführt. Die festgestellten Besonderheiten der Algorithmen sollen im Folgenden näher diskutiert werden.

	ρ_1	ρ_2	ρ_3	h_1	h_2	E
optimales Modell	100.51	12.69	100.35	4.95	1.30	0.769
GAOT, bestes Modell	100.52	16.87	100.36	4.90	1.75	0.769
GAOT, schlechtestes Modell	103.04	62.38	101.02	2.83	12.05	2.149
VFSA, bestes Modell	100.31	13.30	100.30	4.91	1.36	0.800

Tabelle 2: Ergebnisse der Inversion des synthetischen Datensatzes.

4.1 Genetischer Algorithmus

Das Verhalten eines GA wird beeinflusst durch die Codierung der Parameter (logarithmierte Werte der Widerstände und Schichtmächtigkeiten), die Populationsgröße (100 Strings), die Wahl der Selektions-, Kreuzungs- und Mutationsoperatoren, die Kreuzungsrate (90% aller Individuen einer Population werden untereinander gekreuzt), die Mutationsrate (10% aller Individuen mutieren)

sowie die Initialisierung des Zufallszahlengenerators. Es wurden 20 Durchläufe über jeweils 200 Generationen mit unterschiedlich initialisiertem Zufallszahlengenerator berechnet.

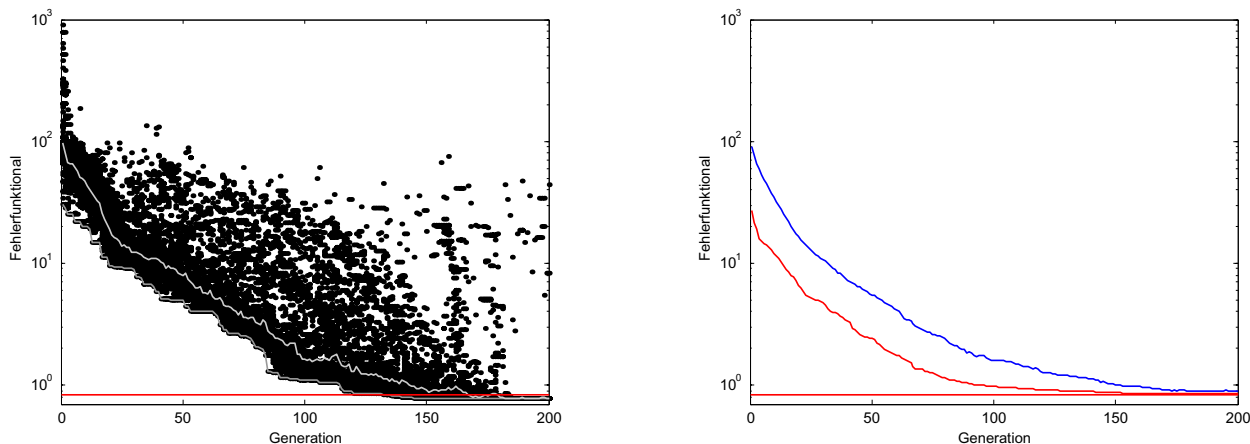


Abbildung 4: Anpassung von Populationen im Genetischen Algorithmus.

Links: Wert des Fehlerfunktionalwertes E für sämtliche 100 Individuen im Verlauf von 200 Generationen. Darüberhinaus sind Mittelwerte, beste Individuen der Generationen und Fehlerschranke von $E = 0.838$ als Linien markiert.

Rechts: Mittelwerte aus 20 Durchläufen mit jeweils unterschiedlich initialisiertem Zufallszahlengenerator. Dargestellt sind die Mittelwerte der Individuen (obere Kurve), die Mittelwerte der bestangepassten Individuen (untere Kurve) sowie o.g. Fehlerschranke.

Die zufallsverteilten Strings der Startpopulation entwickeln sich im Laufe der Evolution zu Strings mit besseren Fehlerfunktionswerten (Abb. 4, links), bis am Ende fast alle Strings um einen Punkt mit minimalem Fehler konzentriert sind. In Abhängigkeit von der Zufallszahlenfolge konvergierte der GA zu unterschiedlich guten Ergebnissen (vgl. Tab. 2), erreichte jedoch im Mittel das zu erwartende Fehlerniveau (Abb. 4, rechts).

4.2 Very Fast Simulated Annealing

Im Vordergrund der Tests stand die Wahl geeigneter Parametersätze für den VFSA-Algorithmus. Es zeigte sich, dass folgende Größen die Konvergenzgeschwindigkeit in erheblichem Maße beeinflussen:

- Intervall dyn der zulässigen Modellparameteränderungen bei konstanter Temperatur
- Anzahl der Züge pro Iterationsschritt
- Starttemperatur T_0 und Abkühlungsrate C
- Initialisierung der Zufallszahlenfolge

Zur Berechnung eines neuen Modells für die Temperatur T wurde folgende Vorschrift verwendet:

$$m_i^{l+1} = \exp(\log(m_i^l) + r_i \cdot dyn_i)$$

Die i logarithmierten Modellparameter wurden im l -ten Zug bei konstanter Temperatur im Intervall $[-dyn_i, +dyn_i]$ mit der Cauchy-verteilten Zufallszahl r_i perturbiert. Für den verwendeten Datensatz zeigte $dyn_i = 0.12, i = 1 \dots N$ die größte Konvergenzgeschwindigkeit. Mit einer geeigneten Wahl von dyn_i lassen sich einzelne Modellparameter unterschiedlich stark perturbieren.

Das beste Modell wurde erreicht mit $T_0 = 1$, $C = 0.995$, $dyn = 0.12$ sowie 10 Zügen pro Iterationsschritt (vgl. Tab. 2).

Das Konvergenzverhalten zeigt typische Eigenschaften eines stochastischen Prozesses mit schwankender Konvergenzgeschwindigkeit. Charakteristisch für die durchgeführten Tests ist das Verharren auf einem hohen Fehlerniveau über längere Zeit (vgl. Abb. 5). Ob dieses Verhalten zufällig ist oder von den verwendeten Parametern abhängt, bleibt weiteren Untersuchungen vorbehalten.

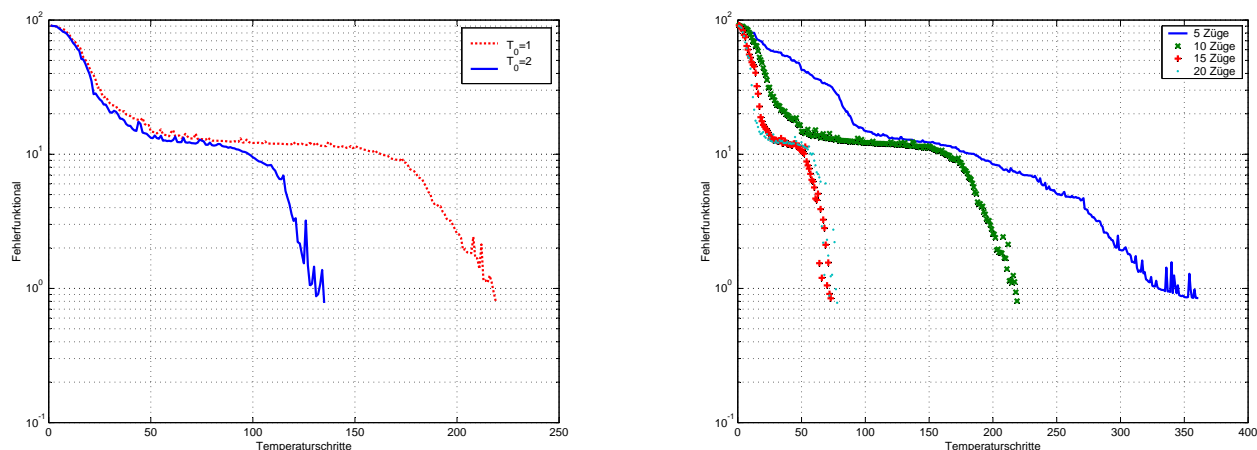


Abbildung 5: Simulated Annealing – Konvergenzverhalten bei verschiedenen Starttemperaturen und 10 Zügen pro Temperaturschritt (links) sowie unterschiedlicher Anzahl von Zügen pro Temperaturschritt (rechts). Der tatsächliche Gesamtrechnenaufwand ergibt sich, wenn man die Anzahl der Temperaturschritte mit der Anzahl der Züge pro Temperatur multipliziert. Im Beispiel rechts führt das zu dem Resultat, dass die Variante "15 Züge" etwa doppelt so schnell konvergiert wie Variante "10 Züge", letztere wiederum langsamer als "5 Züge" und "20 Züge". Ein solches Verhalten ist durch den stochastischen Charakter des Suchprozesses pro Temperaturschritt zu erklären.

5 Zusammenfassung

Die Ergebnisse zeigen, dass Genetische Algorithmen und Simulated Annealing in der Lage sind, auf alternative Weise das Problem der nichtlinearen Inversion zu lösen. Tests mit eindimensionalen Modellen aus der Gleichstromwiderstandsgeoelektrik sind geeignet, prinzipielle Eigenschaften der Algorithmen abzuleiten. Angesichts der rasanten Leistungssteigerung von Computern und der erkennbaren Vorteile nichtlinearer Optimierungsstrategien wird die Anwendung für zwei- und dreidimensionale Inversion rasch in den Blickpunkt des Interesses rücken.

Literatur

- Holland, J. (1975). *Adaption in natural and artificial systems*. The University of Michigan Press, Ann Arbor.
- Houck, C. R., Joines, J. A. und Kay, M. G. (1996). *A genetic algorithm for function optimization: A Matlab implementation*.
- Ingber, L. (1989). *Very fast simulated re-annealing*. Math. Comput. Modelling, 12, 967–973.