

# Variablenreduktion für Surrogat-Modell basierte Optimierung

Frederik Rehbach, Lorenzo Gentile, Thomas Bartz-Beielstein

Institut für Data Science, Engineering, and Analytics, TH Köln  
Steinmüllerallee 1, 51643 Gummersbach  
E-Mail: [firstname.lastname@th-koeln.de](mailto:firstname.lastname@th-koeln.de)

## Kurzfassung

Eine Standardmethode zur effizienten Optimierung von rechenintensiven Problemen ist die *Surrogat-Modell basierte Optimierung* (SMBO). Allerdings steigt bei vielen Surrogat Modellen mit zunehmender Dimensionalität der Rechenaufwand stark an. SMBO ist daher oft nur sehr ineffizient auf hochdimensionalen Problemen anwendbar. In diesem Beitrag wird der Einsatz von Regularisierungsverfahren (Shrinkage) untersucht, um die Dimensionalität des Suchraums zu reduzieren. Wir schlagen einen Hybrid-Algorithmus vor: Regularisierte-Surrogat-Optimierung (RSO). Experimente, die auf künstlichen Testfunktionen und industriellen Anwendungen durchgeführt wurden, zeigen, dass der RSO-Ansatz die Dimensionalität der Modelle erfolgreich reduziert und somit erheblich weniger Rechenzeit benötigt. Gleichzeitig liefert der Algorithmus Ergebnisse, die eine ähnlich hohe Güte aufweisen wie das nicht reduzierte Modell.

## 1 Einführung

Industrielle Probleme besitzen Charakteristiken, die sie von künstlich erzeugten Testfunktionen unterscheiden. Unter anderem sind sie oft hoch-dimensional und teuer zu evaluieren. Ein Standardverfahren für die Optimierung von teuren

Problemen ist Surrogat-Modell basierte Optimierung (SMBO). Eine detaillierte Erklärung zum SMBO Verfahren ist in [1] gegeben.

Im Rahmen laufender Kooperationen mit Industriepartnern reicht es meist nicht einfach nur eine einzelne sehr gute Lösung für ein Problem zu geben. Um auf Akzeptanz in der Industrie zu stoßen, ist es vielmehr notwendig ein Prozessverständnis entwickeln zu können, welches diese Lösung bestätigt und begründet. In dieser Arbeit werden daher zwei Forschungsfragen betrachtet:

**(Q-1):** Wie kann die Interpretierbarkeit von SMBO in industriellen Anwendungen erhöht werden, um offenere Akzeptanz zu erreichen?

**(Q-2):** Wie kann SMBO effizient auf hochdimensionalen Problemen eingesetzt werden?

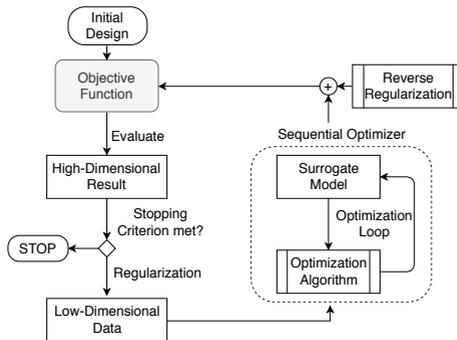


Bild 1: Ablaufdiagramm des RSO Algorithmus

## 2 Regularisierte-Surrogat-Optimierung

Zur effizienten Nutzung von SMBO auf hochdimensionalen Problemen schlagen wir einen zweischrittigen Hybrid-Algorithmus vor: Regularisierte-Surrogat-Optimierung, siehe Abb. 1. Ein initiales Design wird auf der Zielfunktion ausgewertet. In jeder Iteration werden im Regularisierungsschritt die wichtigsten Features des hochdimensionalen Optimierungsproblems erkannt und die Daten für den Optimierungsschritt reduziert. Der Optimierer schlägt

ein auf dem Surrogat gefundenes Optimum als neuen Punkt zur Evaluierung vor. Die fehlenden Features des Punktes werden aufgefüllt, um wieder die ursprüngliche Dimensionalität zu erlangen. Der neue Punkt wird auf der Zielfunktion ausgewertet und die nächste Iteration des Algorithmus beginnt. Es wurden verschiedene Spacefilling-Verfahren getestet, ein Auffüllen mit Zufallszahlen erzeugte hierbei aber die besten Ergebnisse. Als Regularisierungsmethoden werden LASSO (Least Absolute Shrinkage and Selection Operator) [7] und Random Forests (RF) [8] verglichen. RSO ist anwendbar mit verschiedenen Regularisierungsmethoden und Surrogat Modellen. In dieser Arbeit kam Kriging als Surrogat zum Einsatz.

### 3 Experimente

Das erste Industrie-Problem beschäftigt sich mit der Optimierung von elektrostatischen Staubabscheidern (englisch: electrostatic precipitator - ESP), siehe [6]. Als zweite Testfunktion betrachten wir das Sandwich-Structured Composite Plate Design Problem (SCPD). Ein Sandwich Panel ist ein Verbundwerkstoff der aus mehreren Schichten verschiedener Materialien zusammengesetzt wird. Durch die Auswahl der jeweiligen Materialien und deren Eigenschaften wird das Verhalten des Werkstoffes in einer Computersimulation verändert. Ziel der Optimierung ist ein Material zu erzeugen, dass möglichst viel Kraft aufnehmen kann.

Für eine tiefergehende statistische Analyse wurde zusätzlich eine Reihe von künstlichen Testfunktionen verwendet. In allen Testfunktionen ist ein Teil der Variablen wichtig, ein anderer Teil unwichtig für die Optimierung. Tabelle 1 zeigt eine Übersicht der Testfunktionen.

Als Vergleich zu RSO wird eine Standard Implementation von SMBO mit Kriging herangezogen. Da auf den teuren Industrieproblemen nur sehr wenige Evaluierungen möglich sind, erhalten alle Algorithmen ein Budget von 200 Evaluierungen pro Lauf. Zur statistischen Auswertung werden alle Algorithmen je 20 mal wiederholt.

Tabelle 1: Übersicht über alle Testfunktionen und deren Dimensionalität. Angegeben ist die Dimensionalität des Problems, sowie die Anzahl der wichtigen, weniger wichtigen sowie komplett irrelevanten Variablen. Für das ESP Problem sind die wichtigen Variablen nicht bekannt.

Problem	Dimensionen	wichtig	relevant	irrelevant
linketal06dec [2]	10	2	4	4
linketal06sin [2]	10	2	0	8
moon10hd [3]	20	5	10	5
moon10hdc1 [3]	20	5	0	15
moon10hdc2 [3]	20	5	0	15
moon10hdc3 [3]	20	5	0	15
oakoh04 [5]	15	5	5	5
morretal06 [4]	30	2	0	28
ESP-Problem	49	?	?	?
SCPD-Problem	19	2	4	13

## 4 Ergebnisse und Diskussion

Die Ergebnisse der Algorithmen auf den beiden Industrieproblemen sind in Bild 2a und 2b veranschaulicht. Die Ergebnisse zeigen, dass auf beiden Problemen ein Wechsel zu RSO keinen Qualitätsverlust der gefundenen Lösung verursacht. Gleichzeitig ist aber insbesondere auf dem SCPD-Problem zu sehen, dass die nötige Rechenzeit für den Algorithmus drastisch reduziert werden konnte. Die RSO-RF Variante scheint eine striktere Feature-Selektion vorzunehmen als LASSO. Daher wird in dieser Variante auch weniger Rechenzeit benötigt.

Die Ergebnisse aller Experimente sind in Tabelle 2 zusammengefasst. Obwohl auf den zwei Industrieproblemen kein Qualitätsunterschied der gefundenen Lösungen ersichtlich war, ist dies auf manchen künstlichen Testfunktionen der Fall. Unter anderem haben diese Funktionen zu komplexe Zusammenhänge zwischen den Variablen, sodass wichtige und irrelevante Variablen nicht korrekt voneinander getrennt werden können. Insbesondere die RF Variante, die mehr Variablen ausschließt, ist daher von einem Qualitätsverlust betrof-

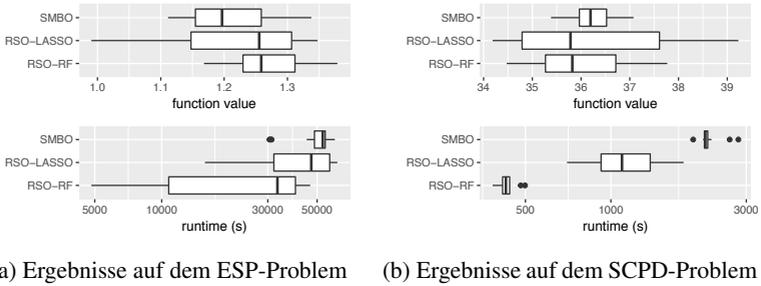


Bild 2: Der jeweils obere Boxplot zeigt die Qualität der besten gefunden Lösung der Optimierungsläufe mit den links beschriebenen Algorithmen. Niedrigere Werte sind besser. Der untere Plot zeigt die vom jeweiligen Algorithmus benötigte Rechenzeit in Sekunden.

Tabelle 2: Übersicht der Ergebnisse aller Algorithmen auf allen Testfunktionen. Die aufgezeigten Zahlen ergeben sich aus einem rangbasierten, paarweisen Test. Der beste Rang (1) ist jeweils fettgedruckt.

	ESP	SCPD	linketal06dec	linketal06sin	moon10hd	moon10hdc1	moon10hdc2	moon10hdc3	oakoh04	morretal06	Mean-Rank
SMBO	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
RSO-LASSO	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	2	<b>1</b>	2	2	2	1.4
RSO-RF	<b>1</b>	<b>1</b>	2	<b>1</b>	2	2	2	2	2	2	1.7

fen. Dennoch ist auf allen Funktionen eine deutliche Reduktion der Algorithmuslaufzeit, ähnlich der Ergebnisse vom SCPD-Problem zu messen. Abschließend möchten wir wieder zu den anfangs definierten Forschungsfragen zurückkehren.

(Q-1): In der Industrie ist es oft sehr schwer komplexe Surrogat Modelle umzusetzen, da sie zu wenig interpretierbare Ergebnisse liefern und daher keine Akzeptanz finden. Dahingegen sind die durch RSO gebauten niedriger dimensionalen Modelle einfacher zu erklären und spiegeln (im Falle des ESP-Problems) beispielsweise die Vermutungen der Ingenieure wider.

(Q-2): Die Ergebnisse zeigen, dass die benötigte Algorithmuslaufzeit, insbesondere die Rechenzeit zum Modellieren der Surrogate, durch RSO stark reduziert werden kann. Gleichzeitig kann RSO auf den meisten Testproblemen Ergebnisse von äquivalenter Qualität finden. Hinzu kommt, dass RSO auch bis zu einer sehr hohen Dimensionalität skaliert werden kann, wohingegen SMBO mit steigender Dimensionalität schlechtere Ergebnisse liefern wird. Ebenso kann die reduzierte Modellierungsdauer RSO für Grenzfälle anwendbar machen, in denen sich SMBO nicht gelohnt hätte.

## Literatur

- [1] T. Bartz-Beielstein und M. Zaefferer „Model-based methods for continuous and discrete global optimization“. Applied Soft Computing, Volume 55, Pages 154-167, 2017.
- [2] C. Linkletter, D. Bingham, N. Hengartner, D. Higdon and K. Ye, „Variable selection for Gaussian process models in computer experiments“. Technometrics, Volume 48, Pages 478-490, 2006.
- [3] H. Moon „Design and analysis of computer experiments for screening input variables“. PhD Thesis - The Ohio State University, 2010.
- [4] M. Morris, L. Moore and M. McKay „Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs“. Journal of Statistical Planning and Inference, Volume 136, Pages 3203-3220, 2006.
- [5] J. Oakley und A. O’Hagan „Probabilistic sensitivity analysis of complex models: a Bayesian approach“. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Volume 66, Pages 751-769, 2004.
- [6] F. Rehbach, M. Zaefferer, J. Stork und T. Bartz-Beielstein, „Comparison of parallel surrogate-assisted optimization approaches“. Proceedings of the Genetic and Evolutionary Computation Conference 2018.

- [7] R. Tibshirani „Regression shrinkage and selection via the lasso“. Journal of the Royal Statistical Society. Series B (Methodological), 1996, Pages 267-288.
- [8] L. Breiman „Random forests“. Machine learning, Volume 45, Pages 5-32, 2001.

