

Seminar Optimierung - Approximation and fitting

Christian Bretzke

Inhaltsverzeichnis

1	Norm Approximation	3
1.1	Verschiedene Interpretation	3
1.2	Gewichtete NAP	4
1.3	Kleinste Quadrate	4
1.4	Minimax Approximation	4
1.5	Summe der absoluten Residuen	5
2	Penalty-Funktionen	5
2.1	Sensitivität gegenüber großen Abweichungen	7
2.2	Robuste Penaltyfunktionen	7
2.3	Kleine Residuen	8
2.4	Approximationen mit Beschränkungen	8
3	Least-Norm Probleme	9
3.1	Umformulierung zum NAP	9
3.2	Interpretationen	9
3.3	Kleinste Quadrate	9
3.4	Kleinste Strafe	10
3.5	Sparse Lösungen	10

1 Norm Approximation

Das einfachste NAP(Norm-Approximations-Problem):

$$\min \|Ax - b\| \tag{1}$$

Daten: $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$

Eine Lösung des NAP heisst oft eine approximiere Lösung in der Norm $\|\cdot\|$ von $Ax \approx b$. Der Vektor

$$r = Ax - b$$

heisst Residuum. Es gibt immer mindestens eine optimale Lösung. Der Optimalwert ist genau dann 0, wenn $b \in \mathfrak{R}(A)$. Das Problem ist viel interessanter und nützlicher wenn $b \notin \mathfrak{R}(A)$. Da für $m = n$ der Optimalwert bei $x = A^{-1}b$ angenommen wird, können wir also $n < m$ annehmen.

1.1 Verschiedene Interpretation

Interpretation als Approximation

Wenn Ax durch

$$Ax = x_1 a_1 + \dots + x_n a_n$$

mit den Spalten $a_{1..n} \in \mathbb{R}^m$ von A ausgedrückt wird, dann ist leicht zu sehen, dass das Ziel des NAP ist, den Vektor b durch Linearkombination der Spalten von A so gut wie möglich anzunähern, wobei hier die Abweichungen in der Norm $\|\cdot\|$ bemessen werden. Dieses Problem heisst auch Regressionsproblem.

Interpretation als Schätzung

Das Problem wird als Schätzung eines Parameters, auf Grundlage einer ungenauen linearen Messung, betrachtet. Es wird also das Modell

$$y = Ax + v$$

mit $y \in \mathbb{R}^m$ einem gemessenen Vektor, $x \in \mathbb{R}^n$ einem Vektor, bestehend aus Parametern, die geschätzt werden sollen, und $v \in \mathbb{R}^m$ einem unbekanntem Messfehler, der als klein in der Norm $\|\cdot\|$ angenommen wird, betrachtet. Aufgabe ist es, eine möglichst genaue Schätzung für x , bei gegebenem y , abzugeben. Aus der Annahme $x = \hat{x}$ folgt implizit, dass v den Wert $y - A\hat{x}$ besitzt. Wenn also die geringeren Werte von v (nach der Norm bemessen) wahrscheinlicher sind als die großen, so ist die wahrscheinlichste Schätzung für x :

$$\hat{x} = \operatorname{argmin}_z \|Az - y\|$$

Geometrische Interpretation

Wird der Unterraum $\mathcal{A} = \mathfrak{R}(A) \subset \mathbb{R}^m$ und ein Punkt $b \in \mathbb{R}^m$ betrachtet, dann ist jeder Punkt in \mathcal{A} , der am nächsten zu b liegt, gerade eine Projektion des Punktes b auf den Unterraum \mathcal{A} . Das wäre praktisch jeder Punkt, der erfüllt:

$$\begin{aligned} \min \|u - b\| \\ \text{s.t.: } u \in \mathcal{A} \end{aligned}$$

Wenn nun ein beliebiges Element aus \mathcal{A} als $u = Ax$ parametrisiert wird, dann ist ersichtlich, dass Problem (1) äquivalent zur Berechnung einer Projektion von b auf \mathcal{A} ist.

Interpretation des besten Designs

Hierbei sind die n Variablen $x_{1..n}$ die Design-Variablen, deren Werte festgelegt werden müssen. Der Vektor $y = Ax$ liefert m Resultate, von denen wir annehmen, dass sie lineare Funktionen von x sind. b beinhaltet die gewünschten Resultate und gesucht wird x mit $Ax \approx b$. Wird die Qualität des Designs durch $\|r\|$ beschrieben, so ist dies gerade Aufgabe (1).

1.2 Gewichtete NAP

Eine Erweiterung des NAP ist das gewichtete NAP (gNAP)

$$\min \|W(Ax - b)\|$$

mit der Gewichtsmatrix $W \in \mathbb{R}^{m \times m}$. Das gNAP kann als NAP mit den Daten $\tilde{A} = WA$ und $\tilde{b} = Wb$ aufgefasst und damit wieder wie Aufgabe (1) behandelt werden. Alternativ: NAP mit normalen Daten und der gewichteten Norm $\|z\|_W = \|Wz\|$. (W muss hier regulär sein)

1.3 Kleinste Quadrate

Durch Quadrieren der Zielfunktion wird eine äquivalente Aufgabe erhalten, das Kleinste-Quadrate-Approximations-Problem, mit den Summen der Quadrate der Residuen als Zielfunktion:

$$\min(\|Ax - b\|_2^2) = \min(r_1^2 + \dots + r_m^2)$$

Diese Aufgabe kann analytisch gelöst werden, als die konvexe quadratische Funktion

$$f(x) = x^T A^T A x - 2b^T A x + b^T b.$$

Ein Punkt x minimiert f also genau dann, wenn

$$\nabla f(x) = 2A^T A x - 2A^T b = 0$$

Das heißt, genau dann, wenn x die Gleichung

$$A^T A x = A^T b$$

erfüllt, die immer eine Lösung hat. Für linear unabhängige Spalten von A hat das Kleinste-Quadrate-Approximations-Problem also die eindeutige Lösung $x = (A^T A)^{-1} A^T b$.

1.4 Minimax Approximation

Wird die ℓ_∞ -Norm verwandt, heisst das NAP

$$\min \|Ax - b\|_\infty = \min(\max_{i=1..m} |r_i|)$$

Minimax-Approximations-Problem oder Chebyshev-Approximations-Problem, da das maximale Residuum minimiert wird. Es kann als LP geschrieben werden:

$$\min t$$

$$s.t. : -t\mathbf{1} \preceq Ax - b \preceq t\mathbf{1}$$

mit den Variablen $x \in \mathbb{R}^n$ und $t \in \mathbb{R}$.

1.5 Summe der absoluten Residuen

Wird die ℓ_1 -Norm verwendet, so heisst das NAP $\min \|Ax - b\|_1 = \min(|r_1| + \dots + |r_m|)$ Approximationsproblem der Summe der absoluten Residuen (oder eine robuste Schätzfunktion). Dies kann auch als LP geschrieben werden

$$\begin{aligned} \min \mathbf{1}^T t \\ \text{s.t. : } -t \preceq Ax - b \preceq t \end{aligned}$$

mit den Variablen $x \in \mathbb{R}^n$ und $t \in \mathbb{R}^m$.

2 Penalty-Funktionen

Bei der ℓ_p -Norm Approximation und $1 \leq p < \infty$ ist die Zielfunktion der Gestalt $(|r_1|^p + \dots + |r_m|^p)^{1/p}$ Äquivalentes Problem $:|r_1|^p + \dots + |r_m|^p$ Dieses ist eine separable und symmetrische Funktion der Residuen. Insbesondere hängt hierbei die Zielfunktion nur noch von der Verteilung der Amplituden der Residuen ab.

Das Penalty-Funktions-Approximations-Problem(PAP) hat die Form

$$\begin{aligned} \min \phi(r_1) + \dots + \phi(r_m) \\ \text{s.t. : } r = Ax - b \end{aligned} \tag{2}$$

mit $\phi : \mathbb{R} \rightarrow \mathbb{R}$, genannt Residuen-Penalty-Funktion. Unter der Annahme, dass ϕ konvex ist, handelt es sich um eine konvexe Optimierungsaufgabe. In vielen Fällen erfüllt die Funktion ϕ die Eigenschaften Symmetrie, Nichtnegativität und $\phi(0) = 0$.

Interpretation

Für ein gewähltes x resultiert die Approximation Ax von b , die den assoziierten Residuenvektor r besitzt. Eine Penalty-Funktion vergibt eine gewisse Strafe für jede Komponente des Residuums, $\phi(r_i)$. Bei dem PAP wird die von den Residuen erzeugte Gesamtstrafe minimiert. Die Form der Penaltyfunktion hat große Auswirkung auf die Lösung der PAP.

Als Beispiel werden nun die ℓ_1 -Norm und ℓ_2 -Norm verglichen: $\phi_1(u) = |u|$ und $\phi_2(u) = u^2$

$$\text{bei } u < 1 \text{ gilt also } \phi_1(u) > \phi_2(u)$$

$$\text{und für } u \rightarrow 0 \text{ gilt } \phi_1(u) \gg \phi_2(u)$$

Die ℓ_1 -Norm Approximation legt gegenüber der ℓ_2 -Norm Approximation größeres Gewicht auf kleine Residualwerte und ϕ_1 betont große Residualwerte weniger als ϕ_2 .

Beispiel

- Wenn $\phi(u) = |u|^p$ mit $p \geq 1$ gesetzt wird, ist das PAP äquivalent zum ℓ_p -NAP. Insbesondere führt $\phi(u) = u^2$ auf die Kleinste-Quadrate-Norm-Approximation und $\phi(u) = |u|$ erzeugt die ℓ_1 -Norm-Approximation.

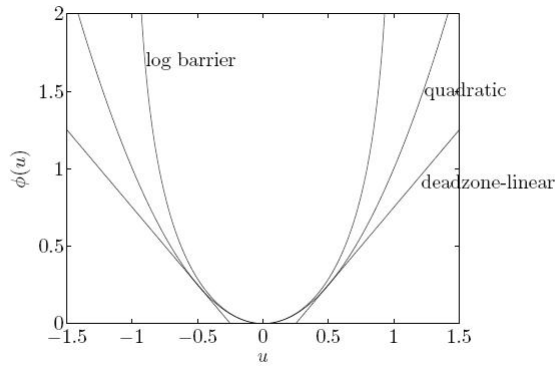
- Die lineare Deadzone-Penalty-Funktion mit $a > 0$ wird beschrieben durch:

$$\phi(u) = \begin{cases} 0 & \text{für } |u| \leq a \\ |u| - a & \text{sonst} \end{cases}$$

- Die logarithmische Barriere-Funktion (mit $a > 0$) hat die Form:

$$\phi(u) = \begin{cases} -a^2 \log(1 - (u/a)^2) & \text{für } |u| < a \\ \infty & \text{sonst} \end{cases}$$

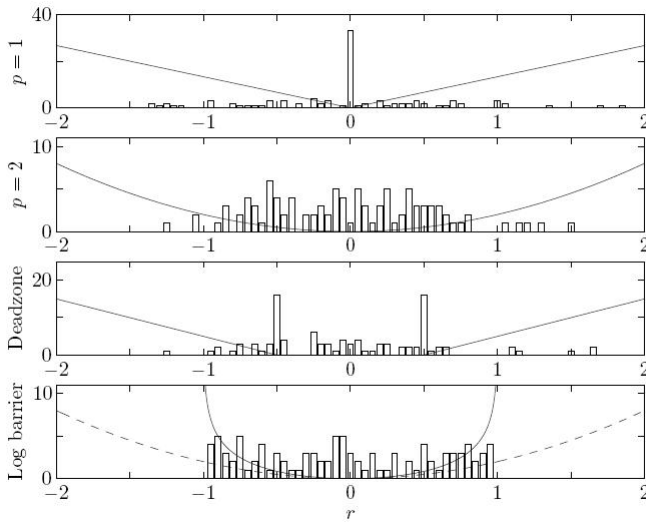
also wird für $u \geq a$ eine unendliche Strafe gegeben.



Beispiel

Zufällig werden Matrix $A \in \mathbb{R}^{100 \times 30}$ und der Vektor $b \in \mathbb{R}^{100}$ gewählt. Es werden die ℓ_1 und ℓ_2 Lösungen für $Ax \approx b$, sowie die Penalty-Funktions-Approximationen mit einer linearen Deadzone-Funktion(a=0.5) und einer logarithmischen Barrierefunktion(a=1) berechnet.

- ℓ_1 -Norm-Penaltyfunktion betont am meisten die kleinen Residuen und am wenigsten die großen.
- ℓ_2 -Norm-Penalty-Funktion gewichtet kleine Residuen nur wenig und große stark.
- Deadzone-Penaltyfunktion legt kein Gewicht auf die Residuen < 0.5 und relativ geringes Gewicht auf große Residuen.
- Die logarithmische Barrierefunktion ist für kleine Residuen sehr ähnlich der ℓ_2 -Norm-Penaltyfunktion, aber legt sehr großes Gewicht auf Residuen > 0.8 und ∞ Gewicht auf Residuen größer 1.



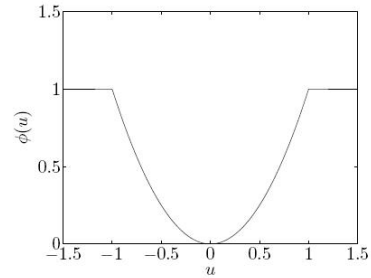
Obige Graphik zeigt die Verteilung der Residuen nach Werten(horizontal) und Häufigkeit des Auftretens(vertikal) zu den verschiedenen Penaltyfunktionen. Man kann erkennen, dass es durchaus relevant ist, welche Penaltyfunktion gewählt wird, da es je nach praktischer Anforderung (z.B. Dünnbesetztheit der Lösung oder einer oberen Schranke für Residuen) gilt, eine richtige Wahl zu treffen.

2.1 Sensitivität gegenüber großen Abweichungen

Ausreißer bedeutet, der Messwert $y_i = a_i^T x + v_i$ hat einen relativ großen Wert v_i . Ausweg:

$$\phi(u) = \begin{cases} u^2 & \text{für } |u| \leq M \\ M^2 & \text{sonst} \end{cases} \quad (3)$$

ϕ legt also eine quadratische Strafe auf alle Werte kleiner M und den festen Wert M^2 auf die größeren, egal wie viel größer das Residuum an dieser Stelle ist. Leider ist die Funktion ϕ nicht konvex und das zugehörige PAP wird zu einem schwierigen, kombinatorischen Optimierungsproblem.



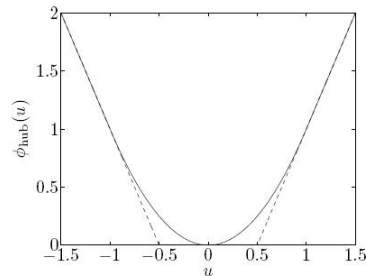
2.2 Robuste Penaltyfunktionen

Ein offensichtliches Beispiel einer robusten Penaltyfunktion ist $\phi(u) = |u|$.

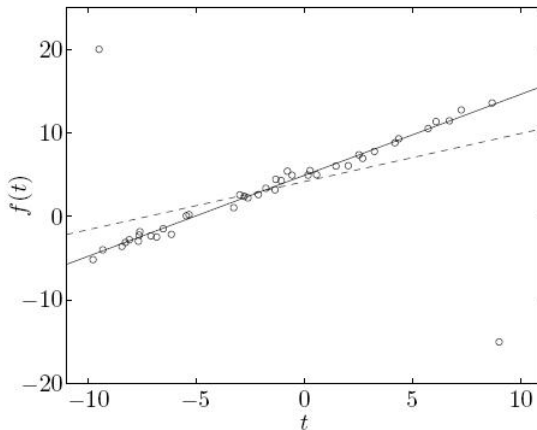
Ein weiteres Bsp. ist die robuste least-squares bzw. Huber-Penaltyfunktion, die durch

$$\phi_{hub}(u) = \begin{cases} u^2 & \text{für } |u| \leq M \\ M(2|u| - M) & \text{sonst} \end{cases} \quad (4)$$

beschrieben wird. Dies entspricht wieder der Kleinsten-Quadrate-Penaltyfunktion für Residuen kleiner M und steigt dann nur noch linear für größere u . Die Huber Penaltyfunktion ist also die zu Funktion (3) nächstmögliche konvexe Funktion.



Beispiel



Die Abbildung zeigt 42 Punkte (t_i, y_i) in einer Ebene, 2 offensichtliche Ausreißer sind am oberen linken und unteren rechten Rand erkennbar. Die gestrichelte Linie zeigt die Kleinst-Quadrate-Approximation der Punkte durch $f(t) = \alpha + \beta t$, mit den Koeffizienten aus der Lösung von $\min \sum_{i=1}^{42} (y_i - \alpha - \beta t_i)^2$ (mit den Variablen α und β). Diese Approximation ist, wie auch im Bild zu erkennen, vom Zentrum der Punkte weg, hin zu den Ausreißern rotiert. Die durchgezogene Linie zeigt die robuste Kleinst-Quadrate-Approximation durch die Huber-Funktion:

$$\min \sum_{i=1}^{42} \phi_{hub}(y_i - \alpha - \beta t_i)$$

mit $M = 1$ und es ist zu sehen, dass diese Approximation um einiges weniger von den Ausreißern gestört wurde.

2.3 Kleine Residuen

Es ist auch möglich, kleine Residuen besonders zu beachten. Penaltyfunktionen, die relativ großes Gewicht auf kleine Residuen legen (z.B.: $|u|$), produzieren tendenziell optimale Residuen, die sehr klein oder 0 sind, was bedeutet, dass wir bei ℓ_1 -Norm-Approximation viele der Gleichungen exakt erfüllt haben, also $a_i^T x = b_i$ für viele i .

Beispiel

Durch hinzufügen von $0 \preceq x$ wird aus (1) die Aufgabe:

$$\begin{aligned} \min & \|Ax - b\| \\ \text{s.t.} & : 0 \preceq x \end{aligned}$$

2.4 Approximationen mit Beschränkungen

Approximationen mit Box-Beschränkungen

Hier fügen wir die Beschränkung $l \preceq x \preceq u$, mit $l, u \in \mathbb{R}^n$ als Parameter ein.

$$\begin{aligned} \min & \|Ax - b\| \\ \text{s.t.} & : l \preceq x \preceq u \end{aligned}$$

Bei einer Schätzung treten Variablenschranken auf, wenn es Vorkenntnisse darüber gibt, in welchen Intervallen die Variablen liegen. Die geometrische Sicht ist die Projektion von b auf das von A erzeugte Bild einer Box.

Approximationen mit Nichtnegativität

$$\begin{aligned} \min & \|Ax - b\| \\ \text{s.t.} & : x \succeq 0, \quad \mathbf{1}^T x = 1 \end{aligned}$$

Diese Aufgabe kann bei Bestimmung von Proportionen oder relativen Häufigkeiten vorkommen, die nichtnegativ und in der Summe 1 sind. (siehe §7.2)

Beschränkungen durch Normumgebungen

$$\begin{aligned} \min & \|Ax - b\| \\ \text{s.t.} & : \|x - x_0\| \leq d \end{aligned}$$

Parameter: x_0 und d

Schätzung: x_0 Vorannahme über x , d ist die größte plausible Abweichung.

Mit $\|x - x_0\|$ kann auch eine trusted region beschrieben werden. Dabei wäre die lineare Relation $y = Ax$ nur eine Approximation einer nichtlinearen Relation $y = f(x)$, die gültig ist, sobald x in der Umgebung von x_0 ist. Also wird nur über die x in der trusted region optimiert.

3 Least-Norm Probleme

(5) Standardproblem:

$$\begin{aligned} \min \|x\| \\ \text{s.t. : } Ax = b \end{aligned}$$

mit den Daten $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, der Variablen $x \in \mathbb{R}^n$ und einer Norm $\|\cdot\|$ im \mathbb{R}^n . Es existiert immer eine Lösung dieser Aufgabe, wenn die linearen Gleichungen $Ax = b$ eine Lösung haben und Aufgabe (5) ist somit eine konvexe Optimierungsaufgabe.

3.1 Umformulierung zum NAP

Aufgabe (5) kann durch Eliminieren der Gleichheitsbeschränkungen als NAP geschrieben werden. Sei x_0 eine beliebige Lösung von $Ax = b$ und $Z \in \mathbb{R}^{n \times k}$ eine Matrix, deren Spalten eine Basis des Kerns von A , $\mathfrak{N}(A)$ bilden. Die allgemeine Lösung von $Ax = b$ kann als $x_0 + Zu$ mit $u \in \mathbb{R}^k$ ausgedrückt werden und somit das Least-Norm-Problem als:

$$\min_u \|x_0 + Zu\|$$

3.2 Interpretationen

Control/Design Interpretation

Wir interpretieren das Problem (5) nun als Aufgabe der optimalen Steuerung oder des optimalen Designs, mit den Variablen $x_{1..n}$, die es zu bestimmen gilt. Bei einer Aufgabe der optimalen Steuerung repräsentieren die Variablen $x_{1..n}$ Eingaben, deren Werte zu wählen sind.

Aus einem gewählten Design x resultieren die m durch den Vektor $y = Ax$ gegebenen, linearen Funktionen. Die m Gleichungen $Ax = b$ repräsentieren m Anforderungen an das Design. Wenn $m < n$, dann ist das Design also unterbestimmt und es existieren $n - m$ Freiheitsgrade (bei $Rg(A) = n$). Von allen möglichen zulässigen Designwahlen wählt Aufgabe (5) also das nach der gewählten Norm kleinste. Dies könnte anschaulich gerade das effizienteste sein, also das kleinst mögliche x , das $Ax = b$ erfüllt.

Geometrische Interpretation

Es gibt auch eine einfache, geometrische Interpretation des Least-Norm-Problems. Die zulässige Menge $x : Ax = b$ ist affin und die Zielfunktion ist der Abstand zwischen x und 0. Das Least-Norm-Problem sucht nun den Punkt der affinen Menge, der am nächsten zu 0 liegt, also sucht es die Projektion der Menge $\{x : Ax = b\}$ auf den Punkt 0.

3.3 Kleinste Quadrate

$$\begin{aligned} \min \|x\|_2^2 \\ \text{s.t. : } Ax = b \end{aligned}$$

ist wieder ein zu (5) äquivalentes Problem und hat eine eindeutige Lösung, die analytisch gefunden werden kann. Es wird die duale Variable $\nu \in \mathbb{R}^m$ eingeführt. Die Optimalitätsbedingungen(KKT)

$$2x^* + A^T \nu^* = 0, \quad Ax^* = b$$

sind bereits erfüllt. Aus der ersten Gleichung wird $x^* = -\frac{1}{2}A^T\nu^*$ und durch Einsetzen in die zweite Gleichung $-\frac{1}{2}AA^T\nu^* = b$ kann geschlossen werden

$$\nu^* = -2(AA^T)^{-1}b, \quad x^* = A^T(AA^T)^{-1}b$$

(bei $\text{Rg}(A) = m < n$ ist $A^T A$ invertierbar)

3.4 Kleinste Strafe

Problem (6):

$$\min \phi(x_1) + \dots + \phi(x_n)$$

$$\text{s.t. : } Ax = b$$

mit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ konvex, nichtnegativ und $\phi(0) = 0$. Der Wert der Penaltyfunktion $\phi(u)$ zeigt wie wenig ein Wert u in einer Komponente von x gewünscht ist. Aufgabe (6) sucht folglich das zulässige x mit der kleinsten Gesamtstrafe.

3.5 Sparse Lösungen

Ausgehend von der Einsicht, dass die ℓ_1 -Norm-Approximation relativ großes Gewicht auf kleine Residuen legt und deshalb viele Residuen sehr klein bzw. 0 sind, tritt auch hier ein ähnlicher Effekt auf.

$$\min \|x\|_1$$

$$\text{s.t. : } Ax = b$$

Das Least- ℓ_1 -Norm-Problem erzeugt also dünnbesetzte Lösungen von $Ax = b$. Zum Finden der Lösungen von $Ax = b$ werden alle Mengen von m Indizes aus $1, \dots, n$ gewählt. Diese sollen gerade ungleich 0 sein. Dies reduziert das Gleichungssystem auf $\hat{A}\hat{x} = b$, mit \hat{A} der $m \times m$ -Untermatrix, die nur aus den gewählten Spalten besteht und $\hat{x} \in \mathbb{R}^m$, dem Subvektor von x , der auch nur die m ausgewählten Komponenten enthält. Für singuläres \hat{A} und $b \notin \mathfrak{R}(\hat{A})$ hat die Gleichung $\hat{A}\hat{x} = b$ keine Lösung, d.h. es gibt kein zulässiges \hat{x} mit $\hat{A}\hat{x} = b$. Für \hat{A} singulär und $b \in \mathfrak{R}(\hat{A})$ existiert eine zulässige Lösung mit weniger als m Komponenten ungleich 0.

Allgemein muss dieses Verfahren allerdings alle $\binom{n}{m}$ möglichen Auswahlen betrachten. Andererseits bietet es eine Methode um dünnbesetzte Lösungen von $Ax = b$ zu finden.