7. Nichtlineare Optimierung

7.1 Einschub - Wiederholung

Sei $A \in \mathbb{R}^{n \times n}$. Dann gelten folgende Definitionen:

- A ist **positiv definit** $\Leftrightarrow \forall x \in \mathbb{R}^n \setminus \{0\}: x^T A x > 0$
- A ist **positiv semidefinit** $\Leftrightarrow \forall x \in \mathbb{R}^n: x^T A x \geq 0$

- **Eigenwert / Eigenvektor**: $\lambda \in \mathbb{C}$ ist Eigenwert zum Eigenvektor $x \in \mathbb{R}^n \setminus \{0\}: \Leftrightarrow Ax = \lambda x$
- A ist positiv semidefinit $\Rightarrow$ alle Eigenwerte sind reell

- Eigenwerte $\lambda_1, \dots, \lambda_n$ (gegebenenfalls mit Vielfachheit), wenn alle Eigenwerte $> 0 \Rightarrow$ Matrix positiv definit
  $\geq 0 \Rightarrow$ Matrix positiv semidefinit

- **Hauptminoren-Kriterium**:

$$\begin{array}{c} \qquad A_1 \quad A_2 \quad A_3 \cdots A_n \\ A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} \end{array}$$

$\rightarrow$ berechne $\det(A_1), \dots, \det(A_n)$

$\hookrightarrow$ wenn alle $> 0 \Rightarrow A$ positiv definit

$\geq 0 \Rightarrow A$ positiv semidefinit

7.2 Einführung

Entwicklung von Algorithmen zur Lösung des unrestringierten Optimierungsproblem (P) „Minimiere“ $f(x)$ über alle $x \in \mathbb{R}^n$ für eine stetige Zielfunktion $f: \mathbb{R} \to \mathbb{R}$ in n Variablen. Da nicht klar ist, ob f auf $\mathbb{R}^n$ einen kleinsten Funktionswert hat, ist (P) äquivalent zu $\inf\{f(x) \mid x \in \mathbb{R}^n\}$. Falls das Infimum für ein $x \in \mathbb{R}^n$ angenommen wird, kann (P) geschrieben werden als $\min\{f(x) \mid x \in \mathbb{R}^n\}$.

Zumeist beschränken wir uns auf Minimierung, da Maximierung umgeschrieben werden kann: $\sup\{f(x) \mid x \in \mathbb{R}^n\} = -\inf\{-f(x) \mid x \in \mathbb{R}^n\}$

Der Wert $\mu := \inf\{f(x) \mid x \in \mathbb{R}^n\}$ ist der Minimalwert von (P). Es sei gesetzt: $\inf\{f(x) \mid x \in \emptyset\} := +\infty$ und $\sup\{f(x) \mid x \in \emptyset\} := -\infty$.

In der Regel ist f nicht-linear, da linear = langweilig. Ist f konvex, so ist (P) ein **konvexes Optimierungsproblem**. Sei $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$.

Def. 7.1 Eine Menge $K \subset \mathbb{R}^n$ heißt **konvex**, wenn gilt $x, y \in K$, $t \in [0,1] \Rightarrow tx + (1-t)y \in K$.

Beispiel

11. Juni 2024

Definition 7.2 Sei $f: \mathbb{R}^n \to \mathbb{R}$ eine Funktion und $K \subset \mathbb{R}^n$ eine konvexe Menge.

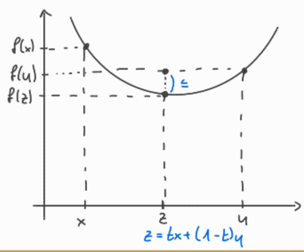
i) f heißt **konvex** auf K, falls gilt: $x, y \in K$, $t \in [0,1] \Rightarrow f(tx + (1-t)y) \leq t\, f(x) + (1-t)\, f(y)$

ii) f heißt **strikt konvex** auf K, falls gilt: $x, y \in K$, $x \neq y$, $t \in (0,1) \Rightarrow f(tx + (1-t)y) < t\, f(x) + (1-t)\, f(y)$

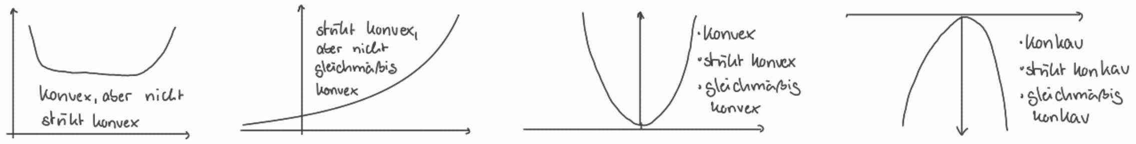
iii) f heißt **gleichmäßig konvex** auf K, falls es eine Konstante $\beta > 0$ (gleichmäßige Konvexitätskonstante) gibt, so dass gilt: $x, y \in K$, $t \in [0,1] \Rightarrow \frac{\beta}{2} \cdot t \cdot (1-t) \cdot \|x-y\|^2 + f(tx + (1-t)y) \leq t \cdot f(x) + (1-t)\, f(y)$.

iv) f heißt **konkav (strikt konkav, gleichmäßig konkav)** auf K, wenn $-f$ konvex (strikt oder gleichmäßig konvex) auf K ist.

Falls $K = \mathbb{R}^n$, dann kann „auf K“ weggelassen werden.



Beispiel



Def. Eine reellwertige Funktion der Form $f(x) := a^T x + \alpha$, $x \in \mathbb{R}^n$ mit $a \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$ heißt **affin** bzw. **affin-linear** und falls $\alpha = 0$ auch nur **linear**.

Glatte konvexe Funktionen werden durch ihre Ableitungen charakterisiert.

Satz 7.3 Sei $K \in \mathbb{R}^n$ konvex und $f \in C^1(K)$ (Ableitung existiert und ist stetig). Dann gilt:

i) f ist auf K genau dann konvex, wenn gilt: $f(y) + \nabla f(y)^T (x-y) \leq f(x) \quad \forall x, y \in K$

ii) f ist auf K genau dann strikt konvex, wenn gilt: $f(y) + \nabla f(y)^T (x-y) < f(x) \quad \forall x, y \in K,\ x \neq y$

iii) f ist auf K genau dann gleichmäßig konvex mit $\beta > 0$, wenn gilt: $\frac{\beta}{2} \|x-y\|^2 + f(y) + \nabla f(y)^T \cdot (x-y) \leq f(x) \quad \forall x, y \in K$

Einschub – Topologie

Sei Menge $K \in \mathbb{R}^n$. Dann gilt:

- K ist abgeschlossen $\iff \forall$ Folgen $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^n$, $\lim_{n\to\infty} x_n = x^* \in \mathbb{R}^n \Rightarrow x^* \in K$
  (Bsp.: $[0,1]$ ist abgeschlossen, $(0,1]$ ist nicht abgeschlossen, Folge $(\frac{1}{n})_{n\in\mathbb{N}}$ mit $\lim_{n\to\infty} \frac{1}{n} = 0$, aber $0 \notin K \Rightarrow$ nicht abgeschlossen)
- $K = \mathbb{Q} \subseteq \mathbb{R}$ ist nicht abgeschlossen
  (Bsp.: $x_n = (1+\frac{1}{n})^n \in \mathbb{Q}$, $\lim_{n\to\infty} (1+\frac{1}{n})^n = e \notin \mathbb{Q}$)
- Sei $U_\varepsilon(x^*) := \{x \in \mathbb{R}^n \mid \|x^* - x\| < \varepsilon\}$ für ein $\varepsilon > 0$ die offene Umgebung von x^*
- Eine Menge K ist offen $\iff \forall x \in K: \exists \varepsilon > 0: U_\varepsilon(x^*) \subset K$
- $K \subseteq \mathbb{R}^n$ kompakt $\iff$ abgeschlossen und beschränkt
- K beschränkt: $\exists M > 0 \ \forall x \in K: U_M(x) \supseteq K$

Satz 1.4 Sei $K \subset \mathbb{R}^n$ konvex und $f \in C^2(K)$ dann gelten:

i) Ist $D^2 f(x)$ für alle $x \in K$ positiv semidefinit, so ist f konvex auf K.

ii) Ist $D^2 f(x)$ für alle $x \in K$ positiv definit, so ist f strikt konvex auf K.

iii) Gibt es eine Konstante $\beta > 0$, so dass $h^T D^2 f(x)\, h \geq \beta \cdot \|h\|^2 \quad \forall h \in \mathbb{R}^n, x \in K$, so ist f gleichmäßig konvex auf K mit der Konstanten β.

iv) Ist K eine offene Menge, so gelten auch die Umkehrungen von i) und iii) (aber nicht ii)).

Definition 1.5

i) x^* heißt globale Lösung von Problem (P), falls $f(x^*) \leq f(x) \ \forall x \in \mathbb{R}$ gilt und strikt globale Lösung, wenn $f(x^*) < f(x) \ \forall x \in \mathbb{R}^n, x \neq x^*$

ii) x^* heißt lokale Lösung von Problem (P), falls ein $\varepsilon > 0$ existiert, so dass $f(x^*) \leq f(x) \ \forall x \in U_\varepsilon(x^*)$ gilt, und strikt lokale Lösung im Fall $f(x^*) < f(x) \ \forall x \in U_\varepsilon(x^*), x \neq x^*$

Man sagt zu x^* auch Minimalpunkt oder Minimierer.

Ist x^* eine lokale bzw. globale Lösung von (P), dann nimmt f ein lokales/globales Minimum in x^* an.

Unterscheide: x^* ist Minimierer von f und $f(x^*)$ ist Minimum von f.

Eine globale Lösung von (P) ist auch stets eine lokale Lösung. Bei konvexen Problemen gilt auch die Umkehrung.

Satz 1.6 Sei $f \in C^0(\mathbb{R}^n)$ konvexe Funktion. Dann gilt:

i) Jede (strikt) lokale Lösung von (P) ist auch (strikt) globale Lösung.

ii) Ist f strikt konvex, dann hat (P) höchstens eine globale Lösung

iii) Die Menge aller globaler Lösungen von (P) ist konvex und abgeschlossen.

Setze $N(x^0) := \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$, genannt Niveaumenge.

Lemma 1.7 Sei $x^0 \in \mathbb{R}^n$ und sei $f \in C^0(\mathbb{R}^n)$. Dann gilt:

i) $x^0 \in N(x^0)$

ii) Ist f konvex, dann ist $N(x^0)$ eine konvexe Menge

iii) $N(x^0)$ ist abgeschlossen.

Satz 1.8 Sei $x^0 \in \mathbb{R}^n$ und $f \in C^0(\mathbb{R}^n)$ ist $N(x^0)$ beschränkt, dann besitzt (P) eine globale Lösung

Für konvexe Optimierungsprobleme $(\tilde{P})$ Minimiere $f(x)$ über alle $x \in Z$ mit konvexer Menge $Z \subset \mathbb{R}^n$ und f gleichmäßig konvex auf Z, trifft folgender Satz eine Aussage über eindeutige Lösungen:

Satz 1.9 Sei $Z \subset \mathbb{R}^n$ nichtleer, abgeschlossen, konvex und $f \in C^1(\mathbb{R}^n)$ auf Z gleichmäßig konvex. Dann gilt:

i) $\forall x^0 \in Z$ ist $\tilde{N}(x^0) := \{x \in Z \mid f(x) \leq f(x^0)\}$ kompakt (abgeschlossen und beschränkt).

ii) $(\tilde{P})$ besitzt genau eine Lösung.

1.2 Positiv definite Matrizen und quadratische Funktionen

13. Juni 2024

Bezeichnen mit $\|\cdot\|$ die Euklidische Vektornorm bzw. die induzierte Matrixnorm die Spektralnorm: $\|A\| := \sup\{\|Ax\| \mid \|x\| = 1\}$. Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ ist genau dann positiv definit, wenn alle ihre Eigenwerte positiv sind. Insbesondere ist diese Matrix nicht-singulär. Bezeichne mit $\lambda_{min}(A)$ und $\lambda_{max}(A)$ ihren kleinsten bzw. größten Eigenwert. Dann ist $\frac{1}{\lambda_{max}(A)}$ bzw. $\frac{1}{\lambda_{min}(A)}$ der kleinste bzw. größte EW von A^{-1}. Somit ist auch A^{-1} symmetrisch und positiv definit. Für die Kondition von A bzgl. der Spektralnorm gilt: $\operatorname{cond}(A) = \|A\| \cdot \|A^{-1}\| = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$

Lemma 7.10

Für $A \in \mathbb{R}^{n \times n}$ symmetrisch gilt:

$$\lambda_{min}(A) \cdot \|x\|^2 \le x^T A x \le \lambda_{max}(A) \cdot \|x\|^2, \; x \in \mathbb{R}^n$$

Ist A zudem positiv definit, gilt:

$$\frac{1}{\lambda_{max}(A)} \cdot \|x\|^2 \le x^T A x \le \frac{1}{\lambda_{min}(A)} \cdot \|x\|^2, \; x \in \mathbb{R}^n$$

Anm. 7.11

Die Ungleichungen in Lemma 7.10 sind scharf, d.h. es gibt ein $x \in \mathbb{R}^n$, für dass sie mit Gleichheit erfüllt sind, bzw. keine besseren Konstanten existieren, die für alle $x \in \mathbb{R}^n$ richtig sind.

Lemma 7.12

Ist $A \in \mathbb{R}^{n \times n}$ symmetrisch, positiv semidefinit und $C \in \mathbb{R}^{m \times n}$, dann ist CAC^T symmetrisch und positiv semidefinit. Ist A positiv definit und Rang(C) = m, dann ist CAC^T positiv definit.

Wichtig in der Optimierung sind quadratische Funktionen. Jede zweimal stetig differenzierbare Funktion lässt sich nach dem Satz von Taylor durch eine solche annähern.

Def 7.13

Eine quadratische Funktion ist eine Funktion $f: \mathbb{R}^n \to \mathbb{R}$, definiert durch $f(x) := \frac{1}{2} x^T \cdot Q \cdot x + c^T x + \alpha$, mit $\alpha \in \mathbb{R}$, $c \in \mathbb{R}^n$ und $Q \in \mathbb{R}^{n \times n}$ symmetrisch. Für die 1. und 2. Ableitung von f gilt: $Df(x) = Qx + c$, $D^2 f(x) = Q$.

Lemma 7.14

Sei $f := \frac{1}{2} x^T Q x + c^T x + \kappa, x \in \mathbb{R}^n$. Dann gilt:

i) f konvex $\iff$ Q positiv semidefinit

ii) f gleichmäßig konvex $\iff$ Q positiv definit

iii) ist f gleichmäßig konvex, so ist $\beta := \lambda_{min}(Q)$ größtmögliche Konvexitätskonstante für f.

7.3 Optimalitätskriterien

Zur Berechnung von Lösungen von (P) ist die Definition der lokalen bzw. globalen Optimalität in der Regel nicht geeignet. Daher interessieren notwendige und hinreichende Optimalitätskriterien für (P).

Satz 7.15

i) Notwendige Optimalitätsbedingung 1. Ordnung:

Sei x^* eine lokale Lösung von (P) und $f \in C^1(\mathbb{R}^n)$. Dann ist $Df(x^*) = 0$.

ii) Notwendige Optimalitätsbedingung 2. Ordnung:

Sei x^* eine lokale Lösung von (P) und $f \in C^2(\mathbb{R}^n)$. Dann gilt: $Df(x^*) = 0$ und $D^2 f(x^*)$ ist positiv semidefinit.

iii) Hinreichende Optimalitätsbedingung 2. Ordnung:

Sei $f \in C^2(\mathbb{R}^n)$ und für x^* gelte $Df(x^*) = 0$ und $D^2 f(x^*)$ ist positiv definit. Dann ist x^* strikt lokale Lösung von (P).

Definition 7.16

Sei $f \in C^1(\mathbb{R}^n)$.

i) Ein $x^* \in \mathbb{R}^n$ mit $Df(x^*) = 0$ heißt stationärer oder kritischer Punkt von f bzw. stationäre oder kritische Lösung von (P).

ii) Ein stationärer Punkt von f, der weder lokaler Minimalpunkt noch lokaler Maximalpunkt von f ist, heißt Sattelpunkt.

Anmerkung

a) Die Bedingung i) ist nicht hinreichend. Nicht jeder stationäre Punkt ist ein Minimum. Für $f(x) = -x^2$, $x \in \mathbb{R}$, ist der einzige stationäre Punkt $x = 0$ ein Minimum. Für $f(x) = x^3$, $x \in \mathbb{R}$, ist der einzige stationäre Punkt ein Sattelpunkt.

b) Die Bedingung ii) ist nicht hinreichend. Für $f(x) = x^3$ ist $Df(x) = f'(0) = 0$ und $D^2 f(x) = f''(0) = 0$. Die Hesse-Matrix ist im stationären Punkt $P = 0$ positiv semidefinit, aber der stationäre Punkt ist weder Minimum noch Maximum.

c) Die Bedingung iii) ist nicht notwendig. Für $f(x) = x^4$, $x \in \mathbb{R}$, ist der stationäre Punkt $x = 0$ ein Minimum, in dem die Hesse-Matrix $D^2 f(0) = f''(0)$ nicht positiv definit ist.

Beispiel 7.17

Untersuche $f(x,y) := x^2 - y^2$ — Gradienten: $Df(x,y) = (2x, -2y)^T$

$g(x,y) := x^3 + y^2$ — $Dg(x,y) = (3x^2, 2y)^T$

$h(x,y) := x^4 + y^2$ — $Dh(x,y) = (4x^3, 2y)^T$

Einziger kritischer Punkt ist immer $(x^*, y^*) = (0,0)^T$

Hesse-Matrizen: $D^2 f(x^*, y^*) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$, $D^2 g(x^*, y^*) = D^2 h(x^*, y^*) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$

Da $D^2 f(x^*, y^*)$ indefinit ist (Eigenwerte sind 2, −2), ist kritische Punkt kein lokales Minimum von f. Durch Anwenden von Satz 7.15 ii) auf $-f$ an, erhält man, dass es auch kein Maximum ist $\Rightarrow (0,0)^T$ ist Sattelpunkt für f.

Da $D^2 g(x^*, y^*) = D^2 h(x^*, y^*)$ nur positiv semidefinit, aber nicht positiv definit, hat Satz 7.15 iii) keine Anwendung (g hat Sattelpunkt, h hat Minimum).

Funktionen können mehrere lokale Minima, Maxima und Sattelpunkte haben. Spezialfall: konvexe Funktionen.

Korollar 7.18

Sei $f \in C^1(\mathbb{R}^n)$ konvex. Dann ist x^* genau dann Lösung von (P), wenn $\nabla f(x^*) = 0$.

Beweis: Für $y := x^*$ folgt aus Satz 7.3 i): $f(x^*) + \nabla f(x^*)^T (x - y) \leq f(x)$ $\forall x \in \mathbb{R}^n$.

Da $\nabla f(x^*) = 0$, folgt $f(x^*) \leq f(x)$ $\forall x \in \mathbb{R}^n$, d.h. x^* ist (globales) Minimum von (P).

Umgekehrt, ist x^* Lösung von (P), so folgt $\nabla f(x^*) = 0$ aus Satz 7.15 i). $\square$

Korollar 7.19

Ist $f \in C^1(\mathbb{R}^n)$ auf $\mathbb{R}^n$ gleichmäßig konvex, so existiert genau ein $x^* \in \mathbb{R}^n$ mit $\nabla f(x^*) = 0$ und x^* ist die eindeutige Lösung von (P).

Beweis: Nach Satz 7.9 für $Z = \mathbb{R}^n$ besitzt (P) genau eine Lösung x^*. Nach Korollar 7.18 ist für diese $\nabla f(x^*) = 0$. $\square$

7.4 Konvergenzrate

Definition 7.20

Sei $(x^k)_k$ eine Folge im $\mathbb{R}^n$ mit $\lim_{k \to \infty} x^k = x^*$.

i) Folge $(x^k)_k$ konvergiert mit (mind.) der Ordnung 1 gegen x^*, wenn $C \in (0,1)$ und $k_0 \in \mathbb{N}$ existieren, so dass $\|x^{k+1} - x^*\| \leq C \cdot \|x^k - x^*\|$, $\forall k \geq k_0$.

ii) Folge $(x^k)_k$ konvergiert mit (mind.) der Ordnung $p > 1$ gegen x^*, wenn $C > 0$ und $k_0 \in \mathbb{N}$, so dass $\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^p$, $\forall k \geq k_0$.

iii) Folge $(x^k)_k$ konvergiert superlinear gegen x^*, wenn es eine Folge $(\varepsilon^k)_k$ mit $\varepsilon_k \geq 0$, $\lim_{k \to \infty} \varepsilon_k = 0$ und ein $k_0 \in \mathbb{N}$ existieren, so dass $\|x^{k+1} - x^*\| \leq \varepsilon_k \cdot \|x^k - x^*\|$, $\forall k \geq k_0$.

Konvergenz (mind.) 1. Ordnung $\rightarrow$ lineare Konvergenz

Konvergenz (mind.) 2. Ordnung $\rightarrow$ quadratische Konvergenz

Äquivalent zur Definition der superlinearen Konvergenz ist: $\lim_{k \to \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0$, falls $x^k \neq x^*$ ab $k \geq k_0$ ist (kann o.B.d.A. angenommen werden).

Quadratische Konvergenz $\Rightarrow$ superlineare Konvergenz $\Rightarrow$ lineare Konvergenz

Lineare Konvergenz mit $C \approx 1$ ist sehr langsam, daher sind Verfahren gewünscht mit $C \ll 1$. Superlineare und quadratische Konvergenz einer Folge im $\mathbb{R}^n$ gelten normunabhängig, die lineare Konvergenz ist normabhängig. Wir betrachten in der Regel die euklidische Norm. Ordnungen in Definition 7.20 heißen Q-Ordnungen. Daneben gibt es noch R-Ordnungen.

Def. 7.21

Sei $(x^k)_k$ Folge im $\mathbb{R}^n$ mit $\lim_{k \to \infty} x^k = x^*$. Dann konvergiert $(x^k)_k$ R-linear (R-superlinear, R-quadratisch) gegen x^*, wenn es eine Folge von $(v_k)_k$ mit $v_k \geq 0$ gibt, welche Q-linear (Q-superlinear, Q-quadratisch) gegen 0 konvergiert und für die $\|x^k - x^*\| \leq v_k$ $\forall k \geq k_0$ gilt.

Q-Konvergenz ist zu „stark", zu „einschränkend". R-Konvergenz erfasst auch dann noch Folgen, die auch hinreichend schnell gegen x^* konvergieren, aber von der Q-Konvergenz nicht erfasst werden.

Beispiel 7.22

1) $a_0 = 1$, $a_1 = \frac{1}{2}$, $a_2 = \frac{1}{4}$, $\dots$, $a_k = \frac{1}{2^k}$

konvergiert Q-linear gegen 0 mit Rate $C = \frac{1}{2}$

2) $b_0 = 1$, $b_1 = 1$, $b_2 = \frac{1}{4}$, $b_3 = \frac{1}{4}$, $b_4 = \frac{1}{16}$, $b_5 = \frac{1}{16}$, $\dots$, $b_k = \frac{1}{4^{\lfloor k/2 \rfloor}}$

konvergiert R-linear gegen 0. Wähle z.B. $v_k = 2a_k$ oder $v_k = a_{k-1}$. $(b_k)_k$ konvergiert nicht Q-linear.

3) $c_0 = \frac{1}{2}$, $c_1 = \frac{1}{4}$, $c_2 = \frac{1}{16}$, $c_3 = \frac{1}{256}$, $\dots$, $c_k = \frac{1}{2^{2^k}}$

konvergiert Q-quadratisch gegen 0 mit $C = 1$. Die Anzahl „gültiger" Stellen in einer Dezimaldarstellung verdoppelt sich nach jedem Folgeglied:

$c_0 = 0{,}5$; $c_1 = 0{,}25$; $c_2 = 0{,}0625$, $c_3 \approx 0{,}003906$; …