Karlsruhe Institute of Technology

Nichtlineare Optimierung
Zusammenfassung WS17/18

Manuel Lang

13. Februar 2018

1 Einführung

1.1 Begriffe

- $\bar{x}$ heißt *lokaler Minimalpunkt* von f auf M, falls eine Umgebung U von $\bar{x}$ mit $\forall x \in U \cap M : f(x) \geq f(\bar{x})$ existiert.
- $\bar{x}$ heißt *globaler Minimalpunkt* von f auf M, falls man obig $U = \mathbb{R}^n$ wählen kann.
- Ein lokaler oder globaler Minimalpunkt heißt *strikt*, falls obig für $x \neq \bar{x}$ sogar die strikte Ungleichung > gilt.
- Zu jedem globalen Minimalpunkt $\bar{x}$ heißt $f(\bar{x})(= v = \min\limits_{x \in M} f(x))$ *globaler Minimalwert*, und zu jedem lokalen Minimalpunkt $\bar{x}$ heißt $f(\bar{x})$ *lokaler Minimalwert*.
- Funktion nicht differenzierbar $\implies$ nichtglattes Optimierungsproblem
- Euklidsche Norm lässt sich durch Quadrieren (Weglassen der Wurzel) zu glattem Problem umformen (optimale Punkte bleiben erhalten)

1.2 Lösbarkeit

- $\alpha \in \mathbb{R}$ wird als *untere Schranke* für f auf M bezeichnet, falls $\forall x \in M : \alpha \leq f(x)$ gilt.
- Das Infimum von f auf M ist die *größte* untere Schranke von f auf M, es gilt also $v = inf_{x \in M} f(x)$ falls $v \leq f(x)$ für alle $x \in M$ gilt (d.h. v ist selbst untere Schranke von f auf M) und $\alpha \leq$ für alle unteren Schranken α von f auf M gilt.

- Definition Lösbarkeit: Das Minimierungsproblem P heißt *lösbar*, falls es ein $\bar{x}$ mit $inf_{x\in M} f(x) = f(\bar{x})$ existiert.
- Satz: *Das Minimierungsproblem P ist genau dann lösbar, wenn es einen globalen Minimalpunkt besitzt.*
- Satz von von Weierstraß: *Die Menge $M \subseteq \mathbb{R}^n$ sei nichtleer und kompakt, und die Funktion $f : M \to \mathbb{R}$ sei stetig. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt und einen globalen Maximalpunkt.*
- Definition untere Niveamenge: Für $\subseteq \mathbb{R}^n, f : X \to \mathbb{R}$ und $\alpha \in \mathbb{R}$ heißt $lev^{\alpha}_{\leq}(f, X) = \{x \in X | f(x) \leq \alpha\}$ *untere Niveaumenge von f auf X zum Niveau α*. Im Fall $X = \mathbb{R}^n$ schreiben wir auch kurz $f^{\alpha}_{\leq} := lev^{\alpha}_{\leq}(f, \mathbb{R}^n)(= \{x \in \mathbb{R}^n | f(x) \leq \alpha\})$.
- Wir führen die Menge der globalen Punkte $S = \{\bar{x} \in M | \forall x \in M : f(x) \geq f(\bar{x})\}$ von P ein.
- Lemma: *Für ein $\alpha \in \mathbb{R}$ sei $lev^{\alpha}_{\leq}(f, M) \neq \emptyset$. Dann gilt $S \subseteq lev^{\alpha}_{\leq}(f, M)$.*
- Verschärfter Satz von Weierstraß: *Für eine (nicht notwendiger beschränkte oder abgeschlossene) Menge $M \subseteq \mathbb{R}^n$ sei $f : M \to \mathbb{R}$ stetig, und mit einem $\alpha \in \mathbb{R}$ set $lev^{\alpha}_{\leq}(f, M)$ nichtleer und kompakt. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt.*
- Korollar (Verschärfter Satz von Weierstraß für unrestringierte Probleme): *Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei stetig, und mit einem $\alpha \in \mathbb{R}$ sei $lev^{\alpha}_{\leq}$ nichtleer und kompakt. Dann besitzt f auf $\mathbb{R}^n$ (mindestens) einen globalen Minimalpunkt.*
- Definition Koerzivität: Gegeben sei eine abgeschlossene Menge $X \subseteq \mathbb{R}^n$ und eine Funktion $f : X \to \mathbb{R}$. Falls für alle Folgen $(x^k) \subseteq X$ mit $lim_k ||x^k|| = +\infty$ auch $lim_k f(x^k) = +\infty$ gilt, dann heißt f *koerziv* auf X.
- Lemma: *Die Funktion $f : X \to \mathbb{R}$ sei stetig und koerziv auf der (nicht notwendigerweise beschränkten) abgeschlossenen Menge $X \subseteq \mathbb{R}^n$. Dann ist die Menge $lev^{\alpha}_{\leq}(f, X)$ für jedes Niveau $\alpha \in \mathbb{R}$ kompakt.*
- Korollar: *Es sei M nichtleer und abgeschlossen, aber nicht allsectionsfont beschränkt. Fernser sei die Funktion $f : M \to \mathbb{R}$ stetig und koerziv auf M. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt.*

1.3 Rechenregeln und Umformungen

- Skalare Vielfache und Summen
  - $\forall \alpha \geq 0, \beta \in \mathbb{R}$: $min_{x\in M}(\alpha f(x) + \beta) = \alpha(min_{x\in M} f(x)) + \beta$.
  - $\forall \alpha \leq 0, \beta \in \mathbb{R}$: $min_{x\in M}(\alpha f(x) + \beta) = \alpha(max_{x\in M} f(x)) + \beta$.
  - $min_{x\in M}(f(x) + g(x)) \geq min_{x\in M} f(x) + min_{x\in M} g(x)$.
  - In obiger Ungleichung kann der strikte Fall > auftreten.

- In den ersten beiden Zeilen stimmen die lokalen bzw. globalen Optimalpunkte der Optimierungsprobleme überein.

- Separable Zielfunktion auf kartesischem Punkt
  - Es seien $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$, $f : X \to \mathbb{R}$ und $g : Y \to \mathbb{R}$. Dann gilt $\min\limits_{(x,y)\in XxY} (f(x)+g(y)) = \min\limits_{x\in X} f(x) + \min\limits_{y\in Y} g(y)$

- Vertauschung von Minima und Maxima
  - Es seien $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$, $M = XxY$ und $f : M \to \mathbb{R}$ gegeben. Dann gilt:
  - $min_{(x,y)\in M} f(x,y) = min_{x\ inX} min_{y\in Y} f(x,y) = min_{y\in Y} min_{x\in X} f(x,y)$.
  - $max_{(x,y)\in M} f(x,y) = max_{x\ inX} max_{y\in Y} f(x,y) = max_{y\in Y} max_{x\in X} f(x,y)$.
  - $min_{x\in X} max_{y\in Y} f(x,y) \geq max_{y\in Y} min_{x\in X} f(x,y)$.
  - In obiger Ungleichung kann der strikte Fall > auftreten.

- Monotone Transformation: Zu $M \subseteq \mathbb{R}^n$ und einer Funktion $f : M \to Y$ mit $Y \subseteq \mathbb{R}$ sei $\psi : Y \to \mathbb{R}$ eine streng monoton wachsende Funktion. Dann gilt $\min\limits_{x\in M} \psi(f(x)) = \psi(\min\limits_{x\in M} f(x))$, und die lokalen bzw. globalen Minimalpunkte stimmen überein.

- Epigrahumformulierung: Gegeben seien $M \subseteq \mathbb{R}^n$ und eine Funktion $f : M\mathbb{R}$. Dann sind die Probleme $P : \min\limits_{x\in\mathbb{R}^n} f(x)$ s.t. $x \in M$ und $P_{epi} : \min\limits_{x,\alpha\in\mathbb{R}^n x\mathbb{R}} \alpha$ s.t. $f(x) \leq \alpha, x \in M$ in folgendem Sinne äquivalent.
  - Für jeden lokalen bzw. globalen Minimalpunkt x^* von P ist $(x^*, f(x^*))$ lokaler bzw. globaler Minimalpunkt von P_{epi}.
  - Für jeden lokalen bzw. globalen Minimalpunkt (x^*, α^*) von P_{epi} ist x^* lokaler bzw. globaler Minimalpunkt von P.
  - Die Minimalwerte von P und P_{epi} stimmen überein.

- Definition Parallelprojektion: Es sei $M \subseteq \mathbb{R}^n x\mathbb{R}^m$. Dann heißt $pr_x M = \{x \in \mathbb{R}^n | \exists y \in \mathbb{R}^m : (x,y) \in M\}$ Parallelprojektion von M auf den „x-Raum “$\mathbb{R}^n$.

- Projektionsumformulierung: Gegeben seien $M \subseteq \mathbb{R}^n x\mathbb{R}^m$ und eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$, die nicht von den Variablen aus $\mathbb{R}^m$ abhängt. Dann sind die Probleme $P : \min\limits_{(x,y)\in\mathbb{R}^n x\mathbb{R}^m} f(x)$ s.t. $(x,y) \in M$ und $P_{proj} : \min\limits_{x\in\mathbb{R}^n} f(x)$ s.t. $x \in pr_x M$ in folgendem Sinne äquivalent:
  - Für jeden lokalen bzw. globalen Minimalpunkt (x^*, y^*) von P ist x^* lokaler bzw. globaler Minimalpunkt von P_{proj}.
  - Für jeden lokalen bzw. globalen Minimalpunkt x^* von P_{proj} existiert ein $y^* \in \mathbb{R}^n$, sodass (x^*, y^*) lokaler bzw. globaler Minimalpunkt von P ist.
  - Die Minimalwerte von P und P_{proj} stimmen überein.

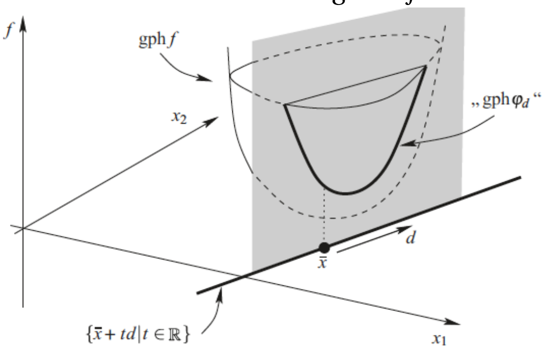
2 Unrestringierte Optimierung

2.1 Optimalitätsbedingungen

2.1.1 Abstiegsrichtungen

- Definition Abstiegsrichtung: Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $\bar{x} \in \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heißt *Abstiegsrichtung* für f in $\bar{x}$ falls $\exists \check{t} > 0 \forall t \in (0, \check{t}) : f(\bar{x} + td) < f(\bar{x})$ gilt.

- Definition eindimensionale Einschränkung: Gegeben seien $f : \mathbb{R}^n \to \mathbb{R}$, ein Punkt $\bar{x} \in \mathbb{R}^n$ und ein Richtungsvektor $d \in \mathbb{R}^n$. Die Funktion $\phi_d : \mathbb{R}^1 \to \mathbb{R}^1, t \mapsto f(\bar{x} + td)$ heißt *eindimensionale Einschränkung* von f auf die durch $\bar{x}$ in Richtung d verlaufende Gerade.



2.1.2 Optimimalitätsbedingung erster Ordnung

- Definition einseitige Richtungsbleitung: Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt an $\bar{x} \in \mathbb{R}^n$ in eine Richtung $d \in \mathbb{R}^n$ *einseitig richtungsdifferenzierbar*, wenn der Grenzwert $f'(\bar{x}, d) := \lim\limits_{t \searrow 0} \frac{f(\bar{x}+td) - f(\bar{x})}{t}$ existiert. Der Wert $f'(\bar{x}, d)$ heißt dann *einseitige Richtungsableitung*. Die Funktion f heißt an $\bar{x}$ *einseitig richtungsdifferenzierbar*, wenn f an $\bar{x}$ in jede Richtung $d \in \mathbb{R}^n$ einseitung richtungsdifferenzierbar ist, und f heißt *einseitig richtungsdifferenzierbar*, wenn f an jedem $\bar{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar ist.

- Lemma: Die Funktion $f : \mathbb{R}^n \to R$ sei an $\bar{x} \in \mathbb{R}^n$ in Richtung $d \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar mit $f'(\bar{x}, d) < 0$. Dann ist d Abstiegsrichtung für f in $\bar{x}$.

- Lemma: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei an einem lokalen Minimalpunkt $\bar{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar. Dann gilt $f'(\bar{x}, d) \geq 0$ für jede Richtung $d \in \mathbb{R}^n$.

- Definition Abstiegsrichtung erster Ordnung: Für eine am Punkt $\bar{x} \in \mathbb{R}^n$ in Richtung $d \in \mathbb{R}^n$ einseitig richtungsdifferenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt d *Abstiegsrichtung erster Ordnung*, falls $f'(\bar{x}, d) < 0$ gilt.

- Definition stationärer Punkt: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei an $\bar{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar. Dann heißt $\bar{x}$ *stationärer Punkt* von f, falls $f'(\bar{x}, d) \geq 0$ für jede Richtung $d \in \mathbb{R}^n$ gilt.

- Als *erste Ableitung* einer partiell differenzierbaren Funktion $f: \mathbb{R}^n \to \mathbb{R}$ an $\bar{x}$ betrachtet man den Zeilenvektor $Df(\bar{x}) := (\partial_{x_1} f(\bar{x}), ..., \partial_{x_n} f(\bar{x}))$ oder auch sein Transponiertes $\nabla f(\bar{x}) := (Df(\bar{x}))^T$.

- Für eine vektorwertige Funktion $f: \mathbb{R}^n \to \mathbb{R}^m$ mit partiell differenzierbaren Komponenten $f_1, ..., f_m$ definiert man die erste Ableitung als $Df(\bar{x}) := \begin{pmatrix} Df_1(\bar{x}) \\ \vdots \\ Df_m(\bar{x}) \end{pmatrix}$. Diese (m, n)-Matrix heißt *Jacobi-Matrix* oder *Funktionalmatrix* von f an $\bar{x}$.

- Satz Kettenregel: Es seien $g: \mathbb{R}^n \to \mathbb{R}^m$ differenzierbar an $\bar{x} \in \mathbb{R}^n$ und $f: \mathbb{R}^m \to \mathbb{R}^k$ differenzierbar an $g(\bar{x}) \in \mathbb{R}^m$. Dann ist $f \circ g: \mathbb{R}^n \to \mathbb{R}^k$ differenzierbar an $\bar{x}$ mit $D(f \circ g)(\bar{x}) = Df(g(\bar{x})) \cdot Dg(\bar{x})$.

- Bei der Anwendung der Kettenregel auf die Funktion $\phi_d(t) = f(\bar{x} + td)$ gilt $k = m = 1$ und $g(t) = \bar{x} + td$. Als Jacobik-Matrix von g erhält man $Dg(t) = d$ und damit $\phi'_d(0) = Df(\bar{x})d$. Das Matrixprodukt aus der Kettenregel wird in diesem Spezialfall also zum Produkt des Zeilenvektors $Df(\bar{x})$ mit dem Spaltenvektor d. Für zwei allgemeine (Spalten-) Vektoren $a, b \in \mathbb{R}^n$ nennt man den so definierten Term $a^T b = \sum_{i=1}^{n} a_i b_i$ auch (Standard-) Skalarprodukt von a und b. Eine alternative Schreibweise dafür ist $\langle a, b\rangle := a^T b$. Wir erhalten also $\phi'_d(0) = \langle \nabla f(\bar{x}), d\rangle$ und können damit zunächst Lemma 2.1.5 umformulieren.

- Lemma 2.1.10: Die Funktion $f: \mathbb{R}^n \to \mathbb{R}$ sei am Punkt $\bar{x} \in \mathbb{R}^n$ differenzierbar, und für die Richtung $d \in \mathbb{R}^n$ gelte $\langle f(\bar{x}), d\rangle < 0$. Dann ist d Abstiegsrichtung für f in $\bar{x}$.

- Für zwei Vektoren $a, b \in \mathbb{R}^n$ besitzt das Skalarprodukt $\langle a, b\rangle$ neben der algebraischen Definition zu $a^T b$ auch die geometrische Darstellung $\langle a, b\rangle = ||a||_2 \cdot ||b||_2 \cdot cos(\angle(a, b))$.

- Satz 2.1.13 Notwendige Optimalitätsbedingung erster Ordnung - Fermat'sche Regel: Die Funktion $f: \mathbb{R}^n \to \mathbb{R}$ sei differenzierbar an einem lokalen Minimalpunkt $\bar{x} \in \mathbb{R}^n$. Dann gilt $\nabla f(\bar{x}) = 0$.

- Die Fermat'sche Regel wird als *Optimalitätsbedingung erster Ordnung* bezeichnet, da sie von der ersten Ableitung der Funktion f Gebrauch macht. Sie motiviert die folgende Definition.

- Definition kritischer Punkt: Die Funktion $f: \mathbb{R}^n \to \mathbb{R}$ sei an $\bar{x} \in \mathbb{R}^n$ differenzierbar. Dann heißt $\bar{x}$ *kritischer Punkt* von f, wenn $\nabla f(\bar{x})$ gilt.

- In dieser Terminologie ist nach der Fermat'schen Regel jeder lokale Minimalpunkt einer differenzierbaren Funktion notwendingerweise kritischer Punkt.

- Definition Sattelpunkt: Die Funktion $f: \mathbb{R}^n \to \mathbb{R}$ sei an $\bar{x} \in \mathbb{R}^n$ differenzierbar. Dann heißt $\bar{x}$ *Sattelpunkt* von f, falls $\bar{x}$ zwar kritischer Punkt von f, aber weder lokaler Minimal- noch Maximalpunkt ist.

- Da die Fermat'sche lediglich eine notwendige, nicht aber eine hinreichende Bedingung ist, sind kritische Punkt lediglich Kandidaten für Minimalpunkt von f, können aber auch Maximal- oder Sattelpunkten entsprechen.

2.1.3 Geometrische Eigenschaften von Gradienten

- Um die geometrische Interpretation des Gradienten $\nabla f(\bar{x})$ vollzuständig zu verstehen, bringen wir ihn mit der unteren Niveaumenge $f_{\leq}^{f(\bar{x})} = \{x \in \mathbb{R}^n | f(x) \leq f(\bar{x})\}$ in Verbindung. Sie ist für Minimierungsverfahren von grundlegender Bedeutung, da einerseits offensichtlich $\bar{x} \in f_{\leq}^{f(\bar{x})}$ gilt und im Vergleich zu $\bar{x}$ „bessere " Punkte x gerade solche sind, die die strikte Ungleichung $f(x) < f(\bar{x})$ erfüllen.

- Cauchy-Schwarz-Ungleichgung: $-||\nabla f(\bar{x})||_2 = -||\nabla f(\bar{x})||_2 \leq \langle f(\bar{x}), d\rangle \leq ||\nabla f(\bar{x})||_2 \cdot ||d||_2 = ||\nabla f(\bar{x})||_2$

- Die kleinstmögliche Steigung $-||\nabla f(\bar{x})||_2$ wird wegen $\nabla f(\bar{x}) \neq 0$ mit $d = -\frac{\nabla f(\bar{x})}{||\nabla f(\bar{x})||_2}$ realisiert und die größtmögliche $+||\nabla f(\bar{x})||_2$ mit $d = +\frac{\nabla f(\bar{x})}{||\nabla f(\bar{x})||_2}$.

- Insbesondere entspricht die Länge $||\nabla f(\bar{x})||_2$ des Gradienten genau dem größtmöglichen Anstieg der Funktion f von $\bar{x}$ aus, und die Richtung des Gradienten zeigt in die zugehörige Richtung des steilsten Anstiegs.

2.1.4 Optimalitätsbedingungen zweiter Ordnung

- univariat = betrachtete Funktion hängt nur von einer Variablen ab

- Satz 2.1.19 (Entwicklungen erster und zweiter Ordnung per univariatem Satz von Taylor)
  - Es sei $\phi : \mathbb{R} \to \mathbb{R}$ differenzierbar an $\bar{t}$. Dann gilt für alle $t \in \mathbb{R}$: $\phi(t) = \phi(\bar{t}) + \phi'(\bar{t})(t - \bar{t}) + o(|t - \bar{t}|)$, wobei $o(|t - \bar{t}|)$ einen Ausdruck der Form $\omega(t) \cdot |t - \bar{t}|$ mit $lim_{t \to \bar{t}} \omega(t) = w(\bar{t}) = 0$ bezeichnet.
  - Es sei $\phi : \mathbb{R} \to \mathbb{R}$ zweimal differenzierbar an $\bar{t}$. Dann gilt für alle $t \in \mathbb{R}$: $\phi(t) = \phi(\bar{t}) + \phi'(\bar{t})(t - \bar{t}) + \frac{1}{2}\phi''(\bar{t})(t - \bar{t})^2 + o(|t - \bar{t}|^2)$, wobei $o(|t - \bar{t}|^2)$ einen Ausruck der Form $\omega(t) \cdot |t - \bar{t}|^2$ mit $lim_{t \to \bar{t}} \omega(t) = w(\bar{t}) = 0$ bezeichnet.

- Lemma 2.1.20: Für $f : \mathbb{R}^n \to \mathbb{R}$, einen Punkt $\bar{x} \in \mathbb{R}^n$ und eine Richtung $d \in \mathbb{R}^n$ seien $\phi_d'(0) = 0$ und $\phi_d''(0) < 0$. Dann ist d Abstiegsrichtung für f in $\bar{x}$.

- Lemma 2.1.21: Für $f : \mathbb{R}^n \to \mathbb{R}$ sei $\bar{x}$ ein lokaler Minimalpunkt. Dann gilt $\nabla(\bar{x}) = 0$, und jede Richtung $d \in \mathbb{R}^n$ erfüllt $\phi_d''(0) \geq 0$.

- Die (n, n)-Matrix $D^2 f(\bar{x}) := D\nabla f(\bar{x}) =:= \begin{pmatrix} \partial_{x_1}\partial_{x_1} f(\bar{x}) & \cdots & \partial_{x_n}\partial_{x_1} f(\bar{x}) \\ \vdots & & \vdots \\ \partial_{x_1}\partial_{x_n} f(\bar{x}) & \cdots & \partial_{x_n}\partial_{x_n} f(\bar{x}) \end{pmatrix}$ heißt *Hesse-Matrix* von f an $\bar{x}$. Als zweite Ableitung sind in ihr Krümmungsinformationen von f and $\bar{x}$ codiert.

- Lemma 2.1.22: Für $f : \mathbb{R}^n \to \mathbb{R}$, einen Punkt $\bar{x} \in \mathbb{R}^n$ und eine Richtung $d \in \mathbb{R}^n$ seien $\langle \nabla f(\bar{x}), d\rangle = 0$ und $d^T D^2 f(\bar{x}) d < 0$. Dann ist §d§ Abstiegsrichtung für f in $\bar{x}$.

- Definition Abstiegsrichtung zweiter Ordnung: Zu $f : \mathbb{R}^n \to \mathbb{R}$ und $\bar{x} \in \mathbb{R}^n$ heißt jeder Richtungsvektor $d \in \mathbb{R}^n$ mit $\langle \nabla f(\bar{x}), d\rangle = 0$ und $d^T D^2 f(\bar{x}) d < 0$ Abstiegsrichtung zweiter Ordnung für f in $\bar{x}$.

- Satz 2.1.27 Notwendige Optimalitätsbedingung zweiter Ordnung: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei zweimal differenzierbar an einem lokalen Minimalpunkt $\bar{x} \in \mathbb{R}^n$. Dann gilt $\nabla f(\bar{x}) = 0$ und $D^2 f(\bar{x}) \geq 0$.

- Eine symmetrische Matrix ist genau dann positiv semidefinit, wenn ihre sämtlichen Eigenwerte nichtnegativ sind.

- Demnach dürfen wir für jede C^2-Funktion (zwei mal stetig differenzierbar) f die Bedingung $D^2 f(\bar{x}) \geq 0$ verifizieren, indem wir die n Eigenwerte der Matrix $D^2 f(\bar{x})$ berechnen und auf Nichtnegativität überprüfen.

- Positive Definitheit bedeutet, dass alle Eigenwerte von $D^2 f(\bar{x})$ strikt positiv sind.

- Satz 2.1.30 Enticklungen erster und zweiter Ordnung per multivariatem Satz von Taylor
  - Es sei $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar in $\bar{x}$. Dan gilt für alle $x \in \mathbb{R}^n$: $f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + o(||x - \bar{x})||$, wobei $o(||x - \bar{x})||$ einen Ausdruck der Form $\omega(x) \cdot ||x - \bar{x}|| mitlim_{x \to \bar{x}} \omega x = \omega \bar{x} = 0$ bezeichnet.
  - Es sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal differenzierbar in $\bar{x}$. Dann gilt für alle $x \in \mathbb{R}^n$: $f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \frac{1}{2}(x - \bar{x})^T D^2 f(\bar{x})(x - \bar{x}) + o(||x - \bar{x}||^2)$, wobei $o(||x - \bar{x}||^2)$ einen Ausruck der Form $\omega(x) \cdot ||x - \bar{x}||^2$ mit $lim_{x \to \bar{x}} \omega(x) = \omega()\bar{x}) = 0$ bezeichnet.

- Satz 2.1.31 Hinreichende Optimalitätsbedingung zweiter Ordnung: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei an $\bar{x} \in \mathbb{R}$ zweimal differenzierbar, und es gelte $\nabla f(\bar{x}) = 0$ und $D^2 f(\bar{x}) \succ 0$. Dann ist $\bar{x}$ ein strikter lokaler Minimalpunkt von f.

- Definition 2.1.35 Nichtdegenerierte kritische und Minimalpunkte: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei an $\bar{x}$ zweimal differenzierbar mit $\nabla f(\bar{x}) = 0$. Dann heißt $\bar{x}$
  - nichtdegenerierter kritischer Punkt, falls $D^2 f(\bar{x})$ nichsingulär ist,
  - nichtdegenerierter lokaler Minimalpunkt, falls $\bar{x}$ lokaler Minimalpunkt und nichtdegenerierter kritischer Punkt ist.

- Lemma 2.1.36: Der Punkt $\bar{x}$ ist genau dann nichtdegenerierter lokaler Minimalpunkt von f, wenn $\nabla f(\bar{x}) = 0$ und $D^2 f(\bar{x}) \succ 0$ gilt.

- Wir definieren $\mathscr{F} = \{f \in C^2(\mathbb{R}^n, \mathbb{R})|$ alle kritischen Punkte von f sind nichtdegeneriert $\}$

- Satz 2.1.37: $\mathscr{F}$ ist C^2_s-offen und -dicht in $C^2(\mathbb{R}^n, \mathbb{R})$.

2.1.5 Konvexe Optimierungsprobleme

- Definition konvexe Menge und Funktionen
  - Eine Menge $X \subseteq \mathbb{R}^n$ heißt konvex, falls $\forall x, y \in X, \lambda \in (0,1) : (1-\lambda)x + \lambda y \; inX$ gilt (d.h. die Verbindungsstrecke von je zwei beliebigen Punkten in X gehört komplett zu X.)
  - Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ heißt eine Funktion $f : X \to \mathbb{R}$ konvex (auf X), falls $\forall x, y \in X, \lambda \in (0,1) : f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y)$ gilt (d.h. der FUnktionsgraph von f verläuft unter jeder seiner Sekanten).
- Satz 2.1.40 C^1-Charakterisierung von Konvexität: Auf einer konvexen Menge $X \subseteq \mathbb{R}^n$ ist eine Funktion $f \in C^1(X, \mathbb{R})$ genau dann konvex, wenn $\forall x, y \in X : f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ gilt.
- Korollar 2.1.41: Die Funktion $f \in C^1(\mathbb{R}^n, \mathbb{R})$ sei konvex. Dann sind die kritischen Punkte von f genau die globalen Minimalpunkt von f.
- Satz 2.1.42 C^2-Charakterisierung von Konvexität: Eine Funktion $f \in C^2(\mathbb{R}^n, \mathbb{R})$ ist genau dann konvex, wenn $\forall x \in \mathbb{R}^n : D^2 f(x) \geq 0$ gilt.