ANDRA DERIVATA-TESTET FÖR KRITISKA PUNKTER, OCH LITE OM KVADRATISKA FORMER

DENNIS ERIKSSON

1. Andagradsderivata testet

Om $f : \mathbb{R}^2 \to R$ var en $\mathcal{C}^1$ funktion, definieriade vi en *kritisk* punkt (P-B: *stationär*) som en punkt (a, b) där $\nabla f(a, b) = 0$. Nedan följer en karaktärisering av dessa kritiska punkter utifrån uppförandet av de kvadratiska termerna i Taylor-utvecklingar. Poängen med detta bevis är att det inte antar att funktionen är $\mathcal{C}^3$, som Persson-Böiers gör, utan antar bara att funktionne är $\mathcal{C}^2$ i en omgivning till punkten. Samtidigt vill jag presentera något som är närmare det bevis de redan gjort själva i boken.
För detta använder vi, som vi bevisade på lektionen, att om f är $\mathcal{C}^2$ i en omgivning till (a, b), så finns en konkret feluppskattning på den linjära approximationen. Detta är samma bevis som det i boken, förutom att vi inte behöver gå lika långt i Taylor-utvecklingen.

$$f(a + h, b + k) = f(a, b) + f_x(a, b)h + f_y(a, b)k + R((a, b),(h, k)) \tag{1}$$

där

$$R((a, b),(h, k)) = f_{xx}(a + \theta h, b + \theta k)h^2 + 2f_{xy}(a + \theta h, b + \theta k)hk + f_{yy}(a + \theta h, b + \theta k)k^2$$

för något $\theta \in (0, 1)$. Eftersom jag tycker detta uttryck är otympligt, och intuitivt inte särskilt tilltalande, skrev jag om detta på matrisform:

Definition 1. *Antag att funktionen f är $\mathcal{C}^2$ i en omgivning till (a, b). Då definierar vi Hessianen till f i (a, b) som*

$$Hess(f)(a, b) = H_f(a, b) = \begin{pmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{pmatrix} (a, b) \tag{2}$$

och dess associerade kvadratisk form som

$$Q_{H_f} := Q_{H_f}(a, b)(h, k) = [h, k] \begin{pmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{pmatrix} [h, k]^t = f_{xx}(a, b)h^2+2f_{xy}(a, b)hk+f_{yy}(a, b)k^2. \tag{3}$$

Här låter vi alltså $(a, b),(h, k)$ vara implicita.

Eftersom jag antar att f är $\mathcal{C}^2$, och inte bara att andra ordningens partiella derivatorna existerar, så är $f_{xy} = f_{yx}$, och matrisen är symmetrisk, dvs. $A = A^T$. Detta är alltså bara ett sätt att packetera informationen från den kvadratiska formen på, som en matris. Man

Date: 27 januari 2016.

säger vidare att Hessianen i en punkt (a, b) är positivt definit, negativt definit, etc, om den associerade kvadratiska formen är det.

Teorem 2. *Låt f vara en $\mathcal{C}^2$ funktion i en omgivning till en kritisk punkt (a, b). Om det då gäller att $H_f(a, b)$ är*

- *positivt definit, så är (a, b) ett strängt lokalt minimum.*
- *negativt definit, så är (a, b) ett strängt lokalt maximum.*
- *indefinit, så är (a, b) en sadelpunkt, dvs. varken lokalt maximum eller lokalt minimum.*

Bevis. Vi visar bara första påståendet, de andra är likartade. Antagandet är alltså att $Q_{H_f}(0, 0)$ är en positivt definit kvadratisk form. Enligt antagande är andra ordningens partiella derivator kontinuerliga, så vi kan skriva om $f_{xx}(a+h, b+k) = f_{xx}(a, b)+R_{xx}(h, k)$ där $R_{xx}(h, k) \to 0$ då $(h, k) \to (0, 0)$. På samma sätt får vi funktioner R_{xy}, R_{yy}. Vi kan då skriva

$$R((a, b),(h, k)) = Q_{H_f}(a+h, b+k)(h, k) = Q_{H_f}(a, b)(h, k)+\left[R_{xx}(h, k)h^2 + 2R_{xy}(h, k)hk + R_{yy}(h, k)k^2\right].$$

Här är $R((a, b),(h, k))$ feltermen i (1) och $Q_{H_f}(0, 0)$ är den kvadratiska formen associerad till Hessianen av f i (a, b). Om nu (h, k) är väldigt litet, kommer termen

$$\left[R_{xx}(h, k)h^2 + 2R_{xy}(h, k)hk + R_{yy}(h, k)k^2\right]$$

gå mot noll även om vi delar med $h^2 + k^2$. Detta följer från uppskattningar av typen $\frac{|hk|}{h^2+k^2} \leq \frac{\sqrt{h^2+k^2}\sqrt{h^2+k^2}}{h^2+k^2} = 1$ och att R_{xx} går mot noll då (h, k) går mot noll. Termen $Q_{H_f}(a, b)(h, k)$, å andra sidan, är neråt uppskattad av en positiv konstant $\delta(h^2 + k^2)$ (detta är ett lemma i boken som föregår satsen). Positiviteten följer från faktumet att Hessianen är positivt definit.

Men detta innebär att vi i princip kan ignorera resttermen i förhållande till den kvadratiska termen $Q_{H_f}(0, 0)$ (denna del av beviset är likadant som boken) och man finner att för små (h, k), så gäller $f(a+h, b+k) = f(a, b)+$positiv term $> f(a, b)$, så att $f(a, b)$ är ett strängt lokalt minimum. □

2. Fler dimensioner

Detta generaliserar till godtyckliga dimensioner, och då är speciellt Hessian-begreppet ett bra sätt att skriva upp informationen på. Detta visar också att Hessianen, dvs. en matris, är bra koncept som fångar upp en rimlig analog till andraderivatan för en funktion av en variabel.

Mer allmänt, givet en $\mathcal{C}^2$-funktion $f : \mathbb{R}^n \to \mathbb{R}$, associerar man till den Hessian-matrisen $H_f = (f_{x_i x_j})_{i,j}$. Det är igen en symmetrisk matris $n \times n$-matris, dvs. den är lika med sitt eget transponat.

Given en kvadratisk $n \times n$-matris $H = (a_{ij})_{i,j}$, associerar man *kvadratisk form* till den, vilket kan definieras som ett uttryck på formen

$$Q(\mathbf{x}) = \mathbf{x}^t A\mathbf{x} = \sum_{i,j} a_{ij}x_i x_j.$$

Om matrisen dessutom är symmetrisk, så är detta samma sak som

$$Q(\mathbf{x}) = \sum a_{ii}x_i^2 + \sum_{i<j} 2a_{ij}x_ix_j.$$

Man definierar analogt positivt definit, negativt definit, etc, för dessa funktioner. I det fallet gäller exakt samma karaktärisering som i satsen i förra stycket, Teorem 2.
En naturlig fråga är således hur man kan verifiera huruvida en symmetrisk matris/kvadratisk form är positivt eller negativt definit. Låt oss betrakta fallet $n = 2$,

$$H = \begin{pmatrix} A & B \\ B & D \end{pmatrix}.$$

Den motsvarande kvadratisk formen är alltså på formen, $Q(h,k) = Ah^2+2Bhk+Ck^2$. Vi vill avgöra när denna funktion alltid är positiv för $(h,k) \neq (0,0)$. Antag för enkelhetens skull att $k \neq 0$. Då är $Q(h,k) > 0$ om och endast om $Q(h,k)/k^2 = A(h/k)^2+2B(h/k)+C$ är positiv. Detta är en kvadratisk funktion i variabeln $t = h/k$. Vi definierar $f(t) = Q(h,k)/k^2 = At^2+2Bt+C$. En sådan funktion är positiv om och endast om f har sitt minimum ovanför noll, dvs. om och endast om f inte har några nollställen och det finns åtminstone ett positivt värde. För att åtminstone ett värde ska vara positivt behöver vi att $A > 0$. Funktions nollställen ges nu också av den vanliga formeln för lösning av kvadratiska ekvationer:

$$f(t) = 0 \Longleftrightarrow t = \frac{-B \pm \sqrt{B^2 - AC}}{A}.$$

Vi ser alltså direkt att det inte finns några nollställen, om och endast om $B^2 - AC < 0$. Men detta är samma sak som att $AC - B^2 = \det H > 0$. Eftersom fallet $h \neq 0$ ger analoga slutsatser, har vi visat:

Proposition 3. *Antag att H är en symmetrisk 2×2-matris. Då är H*

- **+:** *positivt definit om och endast om* $\det H = AC - B^2 > 0, A > 0$.
- **-:** *negativt definit om och endast om* $\det H > 0, A < 0$.

Bevis. Se ovan. Det negativt definita fallet är analogt. □

Följande sats ger en långtgående generalisering av åtminstone vad det innebär att vara positivt definit. För bevis av detta refererar jag till wikipedia. Men jag definierar också, givet en matris A, A_i som den delmatris jag får genom att ta de första i kolonnerna och de första i raderna.

Teorem 4 (Sylvester)**.** *Antag att Q är en kvadratisk form i n variabler, motsvarande en symmetrisk matris $n \times n$-matris A. Då är Q positivt definit om och endast om* $\det A_i > 0$ *för alla del-matriser A_i, $i = 1, \ldots, n$.*