8. april, 2013

# MAT-INF 2360: Obligatorisk oppgave 3

## Innleveringsfrist: 2/5-2013, kl. 14:30

### Informasjon

Den skriftlige besvarelsen skal leveres i obligkassa som står i gangen utenfor ekspedisjonen i 7. et. i Niels Henrik Abels hus senest *kl. 14.30 torsdag 2/5*. Du kan også levere via devilry. Besvarelsen *skal* være skrevet av deg selv, for hånd eller på datamaskin.

Studenter som blir syke eller av andre grunner trenger å søke om utsettelse for denne obligatoriske oppgaven, må ta kontakt med studieadministrasjonen ved Matematisk institutt (7. et. Niels Henrik Abels hus, telefon 22 8558 88, e-post: studieinfo@math.uio.no) i god tid før innleveringsfristen.

Det oppfordres til samarbeid underveis i arbeidet med oppgavene, og gruppelærer og forelesere har anledning til å svare på generelle spørsmål, men kan ikke servere ferdige løsninger. *Den endelige besvarelsen som du leverer skal utarbeides av deg selv, og du må kunne redegjøre for innholdet ved en eventuell muntlig høring (aktuelt ved mistanke om avskrift).*

Husk at de tre obligatoriske oppgavene i MAT-INF 2360 alle må bestås for å kunne gå opp til endelig eksamen i kurset. *For å få bestått på denne tredje obligatoriske oppgaven må du gjøre seriøse løsningsforsøk på alle oppgavene, og minst 6 av de 8 deloppgavene bør være riktig besvart.*

I kapittel 1 i kompendiet om ikkelineær optimering forklarte vi at maximum-likelihood er en av de viktige anvendelsene av ikkelineær optimering. Vi skal se litt mer på hva dette innebærer i praksis. Utgangspunktet er at man har en sannsynlighetstetthetsfunksjon $p(x;\alpha,\beta,\ldots)$ for et eller annet fenomen, der $\alpha,\beta,\ldots$ er ukjente parametre (det kan være mange flere parametre enn det vi har listet opp her).

Vi skal se på en modell fra fysikk for desintregrering av myoner (se Eksempel 1.3 i kompendiet). Vinkelen $\theta$ i elektronstråling ved desintregrering av myoner har en fordeling med sannsynlighetstetthet

$$p(x;\alpha) = \frac{1+\alpha x}{2} \tag{1}$$

for $x \in [-1,1]$, der $x = \cos\theta$, og der $\alpha$ er en ukjent parameter i $[-1,1]$. Målet vårt er å estimere $\alpha$ ved hjelp av $n$ målinger $\mathbf{x} = (x_1,\ldots,x_n)$. I metoden for maximum likelihood gjøres dette ved at vi maksimerer likelihoodfunksjonen $g(\alpha) = \prod_{i=1}^n p(x_i;\alpha)$.

**1 (Oppgave 1.4 i kompendiet).** Forklar hvorfor det å maksimere $g$ er det samme som å maksimere $\ln g$.

Det viser seg at det ofte er enklere å maksimere $\ln g$ enn $g$ selv, siden logaritmen til et produkt er summen av logaritmene. Vi skal derfor jobbe med $\ln g$, som kalles for log-likelihoodfunksjonen. Mer presist skal vi jobbe med funksjonen $f(\alpha) = -\ln g(\alpha)$, siden å finne maksimum til $g$ er det samme som å finne minimum til $f$, og siden kompendiet er formulert for å finne minimum til funksjoner. Vi skal med andre ord minimere funksjonen

$$f(\alpha) = -\ln g(\alpha) = -\ln\left(\prod_{i=1}^n p(x_i;\alpha)\right) = -\sum_{i=1}^n \ln((1+\alpha x_i)/2). \tag{2}$$

**2.** Når vi minimerer $f$ har vi også tilleggsbetingelsene $-1 \leq \alpha \leq 1$. Omformuler problemet slik at det er et betinget minimeringsproblem med to ulikhetsbetingelser $g_1$ og $g_2$.

**3.** Skriv opp KKT-betingelsene for problemet du fant i forrige oppgave, og drøft løsningene du får for følgende muligheter:

1. Kun $g_1$ er aktiv
2. Kun $g_2$ er aktiv
3. Hverken $g_1$ eller $g_2$ er aktive
4. Både $g_1$ og $g_2$ er aktive

I Eksempel 1.3 i kompendiet regnet vi ut $f'(\alpha)$ og $f''(\alpha)$ til å være

$$f'(\alpha) = -\sum_{i=1}^{n} \frac{x_i/2}{(1+\alpha x_i)/2} = -\sum_{i=1}^{n} \frac{x_i}{1+\alpha x_i}$$
$$f''(\alpha) = \sum_{i=1}^{n} \frac{x_i^2}{(1+\alpha x_i)^2}$$

Fra dette fulgte det at $f$ er konveks. Tegner vi et fortegnskjema for $f'(\alpha)$ over hele $\mathbb{R}$ ser vi da at

- $f$ har nøyaktig ett minimum i hvert intervall av formen $[-1/x_i, -1/x_{i+1}]$ når vi lister opp $x_i$ i voksende rekkefølge.
- $f'(\alpha) \to 0$ når $\alpha \to \pm\infty$.
- $f'(\alpha) \to \infty$ når $\alpha \to -1/x_i$ nedenfra.
- $f'(\alpha) \to -\infty$ når $\alpha \to -1/x_i$ ovenfra.

**4.** Forklar ut fra fortegnskjemaet at det vil alltid kun være en kandidat fra mulighet 1 til 3 over som vil være kandidat til minimum for $f$, uavhengig av måleverdiene $x_i$.

Siden $f$ er konveks kan vi bruke resultater fra kompendiet til å begrunne at Newtons metode for å finne eventuelle punkter der $f'(\alpha) = 0$ konvergerer, så lenge vi starter nær nok. Siden vi allerede har regnet ut den deriverte og andrederiverte for $f$ kan vi sette disse direkte inn i Newton's metode for å finne minimum til $f$ (=nullpunkt til $f'$).

**5.** I Kapittel 4 i kompendiet om ikkelineær optimering forklares det hvordan man kan lage en litt mer avansert variant av Newtons metode som fungerer for funksjoner i flere variable, og som bruker det vi kaller *backtracking line search*. I Oppgave 4.12 på s. 49 blir du bedt om å programmere denne. Koden fra denne oppgaven kan du finne i funksjonen `newtonbacktrack`, som ligger på det vanlige området for matlabkode for kurset (http://folk.uio.no/oyvindry/matinf2360/matlab/). Kommenter denne koden, ut fra teorien i Kapittel 4. Kommenter også hvordan du kan endre koden slik at den tar høyde for muligheten at det ikke finnes noe punkt der den deriverte er 0.

**6 (Oppgave 4.10 i kompendiet).** Funksjonen `newtonbacktrack` kaller en annen funksjon, `armijorule`, som implementerer Likning (4.7) i kompendiet (Oppgave 4.11 i kompendiet), og som du også kan finne på det vanlige området for matlabkode for kurset. Hva vil kunne gå galt i denne funksjonen hvis $\nabla^2 f$ er negativ definit (det vil si at alle egenverdiene til $\nabla^2 f$ er negative)? Merk at dette ikke

vil skape problemer for funksjonen $f(\alpha)$ vi bruker her, siden den er konveks (slik at $\nabla^2 f$ er positiv definit)?
**Hint**: Sett inn Taylorapproksimasjonen

$$f(\boldsymbol{x}_k + \beta^m s\boldsymbol{d}_k) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T(\beta^m s\boldsymbol{d}_k)$$

i (4.7), og husk at $\sigma$ der er valgt slik at $\sigma < 1$.

**7 (Oppgave 4.13 a. i kompendiet).** Kjør funksjonen `newtonbacktrack` med $f$, $f'$, og $f''$ som parametre når $n = 10$ og

$$\mathbf{x} = (0.4992, -0.8661, 0.7916, 0.9107, 0.5357, 0.6574, 0.6353, 0.0342, 0.4988, -0.4607)$$

Bruk startverdi $\alpha_0 = 0$ for Newtons metode. Husk at du kan sende inn en funksjon som $f(x) = x^2$ som parameter ved å sende med `@(x)x.^2` som parameter. Hvilket estimat for minimumspunktet til $f$, og dermed til $\alpha$, fikk du?

De ti punktene over ble generert fra en sannsynlighetsfordelig der $\alpha = 0.5$. Svaret du fikk i Oppgave 5 viste seg å være gansk langt fra dette. La oss også se på hvor mange måleverdier vi bør bruke for å få ganske presise estimater på $\alpha$. Du kan bruke funksjonen

```
function ret=randmuon(alpha,m,n)
```

til å generere en $m \times n$-matrise med måleverdier generert etter en sannsynlighetsfordeling med gitt verdi for $\alpha$. Denne funksjonen kan du finne på det vanlige området for matlabkode for kurset.

**8 (Oppgave 4.13 b. i kompendiet).** Med $\alpha = 0.5$, generer $n = 10$ måleverdier ved hjelp av funksjonen `randmuon`, og finn maximum-likelihood-estimatet som over. Gjenta dette ti ganger, og plott de ti estimatene du har fått. Gjenta for $n = 1000$, og for $n = 100000$ (i alle tilfellene skal du altså plotte 10 maximum-likelihood estimater). Hvor mange måleverdier trenger vi for å få en pålitelig maximum-likelihood estimator for $\alpha$?