Class Friday Mar 11 ⇐ list of topics for test

Tues Mar 15

TA Session ~~Thurs 17~~ receive HW ~~cancelled~~

~~Frid 18~~

test Mar 22 **

---

optim method = "Nelder-Mead"

(no derivs. needed
slow)

method = "BFGS" quasi-Newton

recall NR $\underset{\sim}{\hat{\theta}}^{(t+1)} = \underset{\sim}{\hat{\theta}}^{(t)} - \underbrace{\left(\ell''(\underset{\sim}{\hat{\theta}}^{(t)})\right)^{-1}}_{\text{replace w/ something else}} \ell'(\hat{\theta}^{(t)})$

~~$\ell'(\hat{\theta}^{(t+1)})$ for next computation~~

- we have $\ell'(\hat{\theta}^{(t-1)})$ when we start step t

need $\ell'(\hat{\theta}^{(t)})$ " " " $t+1$

$$\begin{pmatrix} \hat{\theta}_1^{(t+1)} \\ \vdots \\ \hat{\theta}_p \end{pmatrix} = \begin{pmatrix} \hat{\theta}_1^{(t)} \\ \vdots \\ \hat{\theta}_p^{(t)} \end{pmatrix} - \left[\quad \right]^{-1} \left[\quad \right]$$

BFGS uses an approx. to $[\ J\ ]^{-1}$ obtained from $\ell'(\hat\theta^{(t-1)})$, $\ell'(\hat\theta^{(t)})$

Let $J_{t+1} = \ell''(\hat\theta^{(t+1)})$

$J_t = \ell''(\hat\theta^{(t)})$

$$J_{t+1} = J_t - \underbrace{\frac{J_t d_t d_t' J_t}{d_t' J_t d_t}}_{\text{rank 1 matrix}} + \underbrace{\frac{g_t g_t'}{d_t' g_t}}_{\text{rank 1 matrix}}$$

$d_t = \hat\theta^{(t+1)} - \hat\theta^{(t)}$

$g_t = \ell'(\hat\theta^{(t+1)}) - \ell'(\hat\theta^{(t)})$

$J_{t+1} - J_t$ has rank at most 2

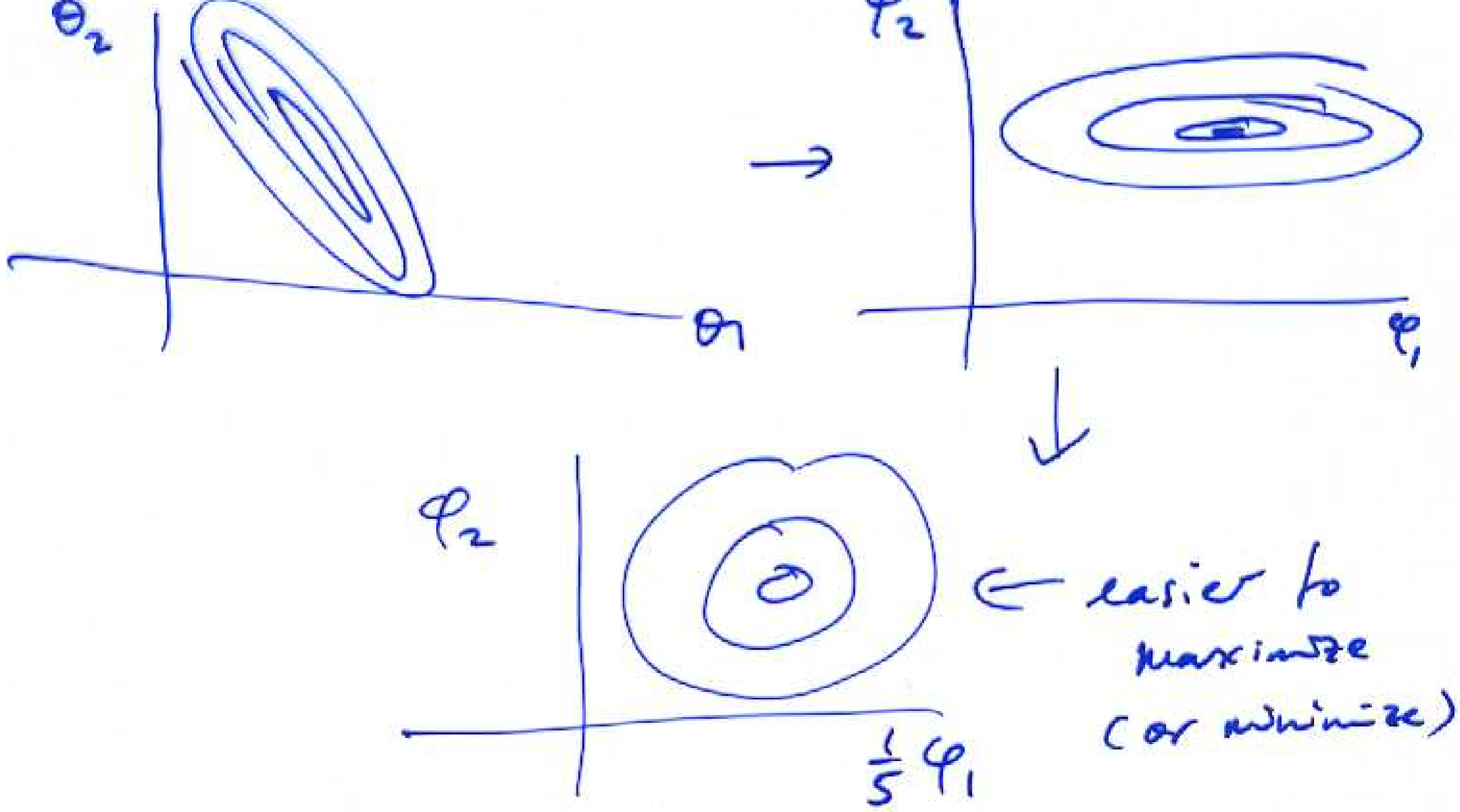
J_t, J_{t+1} have rank p

$\Rightarrow$ J_{t+1}^{-1} can be obtained from J_t in $O(p^2)$ operations — BFGS does this

brute force is $O(p^3)$

Thisted Scientific Computation in Statistics

Functions with long 'ridges' are hard to minimize (or maximize)



- in practice: rescale data so each 'variable' is roughly on same scale

$\ell(\theta_1, \theta_2)$ will have 'nice' contours often,

if $$E \frac{\partial^2 \ell(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2} = 0$$

In $\Gamma(\beta, \mu)$ $$E \frac{\partial^2 \ell(\mu, \beta)}{\partial \mu \partial \beta} = 0$$ orthog.

also $\hat{\mu}_\beta \equiv \hat{\mu}$

$\to$ (✝) $\frac{1}{\Gamma(\beta)} y^{\beta-1} \lambda^{\beta} e^{-\lambda y}$ Gamma$(\beta, \lambda) \leftarrow$ not $\perp$

$\frac{1}{\Gamma(\beta)} \left(\frac{\beta}{\mu}\right)^{\beta} y^{\beta-1} e^{-\frac{\beta}{\mu} y}$ Gamma$(\beta, \mu) \leftarrow$ are <u>orthog.</u>

<u>Thm</u> If $\theta_1 \perp \theta_2$ in this sense then

$$\hat{\theta}_1(\theta_2) = \hat{\theta}_1 + O_p\left(\frac{1}{n}\right)$$

if θ_1 not $\perp$ θ_2 $\hat{\theta}_1(\theta_2) = \hat{\theta}_1 + O_p\left(\frac{1}{\sqrt{n}}\right)$

$\to$ Constrained mle for θ_1 when θ_2 is fixed

Mixture model

$$f(y; p, \mu_1, \mu_2, \sigma_1, \sigma_2)$$

$$= p \frac{1}{\sigma_1} \frac{e^{-\frac{1}{2\sigma_1^2}(y - \mu_1)^2}}{\sqrt{2\pi}} + \frac{(1-p)}{\sqrt{2\pi}} \frac{1}{\sigma_2} e^{-\frac{1}{2\sigma_2^2}(y-\mu_2)^2}$$