

Mathematical Concepts 3

**Solution 1:**

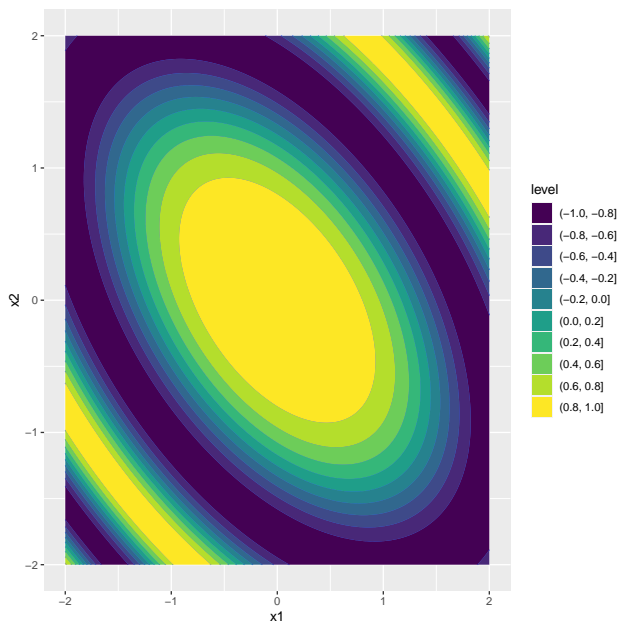
Optimality in 2d

(a) `library(ggplot2)`

```
f <- function(x, y) cos(x^2 + y^2 + x*y)
x = seq(-2, 2, by=0.01)
xx = expand.grid(X1 = x, X2 = x)

fxx = f(xx[,1], xx[,2])
df = data.frame(xx = xx, fxx = fxx)

ggplot(df, aes(x = xx.X1, y = xx.X2, z = fxx)) +
  geom_contour() +
  geom_contour_filled() +
  xlab("x1") +
  ylab("x2")
```



(b)  $\nabla f = (\sin(x_1^2 + x_2^2 + x_1x_2)(2x_1 + x_2), \sin(x_1^2 + x_2^2 + x_1x_2)(2x_2 + x_1))^T$

(c)  $\nabla^2 f = \begin{pmatrix} \cos(u)(2x_1 + x_2)^2 + 2\sin(u) & \cos(u)(2x_1 + x_2)(2x_2 + x_1) + \sin(u) \\ \cos(u)(2x_1 + x_2)(2x_2 + x_1) + \sin(u) & \cos(u)(2x_2 + x_1)^2 + 2\sin(u) \end{pmatrix}$  with  $u = x_1^2 + x_2^2 + x_1x_2$ .

(d) Let  $u : \mathbb{R}^2 \rightarrow \mathbb{R}, (x_1, x_2) \mapsto x_1^2 + x_2^2 + x_1x_2$ .

$$\nabla u = (2x_1 + x_2, x_1 + 2x_2)^T \stackrel{!}{=} \mathbf{0} \iff \mathbf{x} = \mathbf{0}$$

$$\nabla^2 u = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \Rightarrow \mathbf{v}^T \nabla^2 u \mathbf{v} = 2v_1^2 + 2v_1v_2 + 2v_2^2 = v_1^2 + v_2^2 + (v_1 + v_2)^2 \geq 0 \text{ (equality only holds if } \mathbf{v} = \mathbf{0} \text{)}$$

$\Rightarrow \nabla^2 u$  is positive definite.

$$\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^T} - \cos(u) = \cos(u) \left( \frac{\partial}{\partial \mathbf{x}} u \right)^T \frac{\partial}{\partial \mathbf{x}} u + \sin(u) \frac{\partial}{\partial \mathbf{x} \partial \mathbf{x}^T} u$$

Case  $\mathbf{x} = \mathbf{0}$  :  $\left( \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^T} - \cos(u) \right) (\mathbf{0}) = \cos(u) \mathbf{0} + 0 \frac{\partial}{\partial \mathbf{x} \partial \mathbf{x}^T} u = \mathbf{0}$  (is p.s.d.)

Case  $\mathbf{x} \in S_{\mathbb{R}} \setminus \{\mathbf{0}\} : u \in (0, \pi/2) \Rightarrow \sin(u) > 0, \cos(u) > 0$   
 $\mathbf{v}^\top \left( \frac{\partial}{\partial \mathbf{x}} u \right)^\top \frac{\partial}{\partial \mathbf{x}} u \mathbf{v} = z \cdot z > 0$  with  $z = \frac{\partial}{\partial \mathbf{x}} u \mathbf{v} \Rightarrow \left( \frac{\partial}{\partial \mathbf{x}} u \right)^\top \frac{\partial}{\partial \mathbf{x}} u$  is p.s.d.  
 $\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^\top} - \cos(u) = \underbrace{\cos(u) \left( \frac{\partial}{\partial \mathbf{x}} u \right)^\top \frac{\partial}{\partial \mathbf{x}} u}_{\text{p.s.d.}} + \underbrace{\sin(u) \frac{\partial}{\partial \mathbf{x} \partial \mathbf{x}^\top} u}_{\text{p.d.}}$  is p.d.  
 $\Rightarrow f|_{S_{\mathbb{R}}}$  is convex

(e)  $\nabla f|_{S_{\mathbb{R}}} = - \underbrace{\cos(u)}_{>0} \nabla u$  and  $\nabla u = \mathbf{0} \iff \mathbf{x} = \mathbf{0} \stackrel{(e)}{\Rightarrow} \mathbf{0}$  is local minimum

(f)  $f(\mathbf{0}) = -1$  and  $\cos : \mathbb{R} \rightarrow [-1, 1]$ . From this it follows that  $\mathbf{0}$  must be a global minimum of  $f$  since no element of the image of  $f$  is smaller than  $-1$ .

## Solution 2:

Optimality in d dimensions

(a)  $\text{Var}(\mathbf{w}^\top \mathbf{X} - \mathbf{Y}) = \text{Var}(\mathbf{w}^\top \mathbf{X}) + \text{Var}(\mathbf{Y}) - 2\text{Cov}(\mathbf{w}^\top \mathbf{X}, \mathbf{Y}) = \mathbf{w}^\top \Sigma_{\mathbf{X}} \mathbf{w} + \text{Var}(\mathbf{Y}) - 2\mathbf{w}^\top \Sigma_{\mathbf{X}\mathbf{Y}}$ . This is a quadratic form in  $\mathbf{w}$  and  $\Sigma_{\mathbf{X}}$  is p.s.d. (since it is a covariance matrix)  $\Rightarrow f$  is convex.

(b)  $\nabla f = 2\Sigma_{\mathbf{X}} \mathbf{w} - 2\Sigma_{\mathbf{X}\mathbf{Y}}, \nabla^2 f = 2\Sigma_{\mathbf{X}}$

(c)  $\nabla f \stackrel{!}{=} \mathbf{0} \iff 2\Sigma_{\mathbf{X}} \mathbf{w} - 2\Sigma_{\mathbf{X}\mathbf{Y}} = \mathbf{0} \iff \Sigma_{\mathbf{X}} \mathbf{w} = \Sigma_{\mathbf{X}\mathbf{Y}}$ . This system of linear equations has a unique solution if  $\Sigma_{\mathbf{X}}$  is non-singular. If  $\Sigma_{\mathbf{X}}$  is non-singular it follows that  $\mathbf{w} = \Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}}$ . In this case  $\Sigma_{\mathbf{X}}$  is p.d. since no eigenvalue can be zero,  $f$  is strictly convex and the local minimum is global.

(d) First condition: Since  $\mathbf{w}$  exists  $\Sigma_{\mathbf{X}}$  must be non-singular.

Then  $\Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}} = \mathbb{E}((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^\top)^{-1} \mathbb{E}((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^\top)$

Second condition: If  $\mathbb{E}(\mathbf{X}) = \mathbf{0}, \mathbb{E}(\mathbf{Y}) = \mathbf{0}$  then

$\Sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}} = (\mathbb{E}(\mathbf{X}\mathbf{X}^\top))^{-1} \mathbb{E}(\mathbf{X}\mathbf{Y}^\top).$

$n(\mathbf{x}_{1:n}^\top \mathbf{x}_{1:n})^{-1}$  is a consistent estimator of  $(\mathbb{E}(\mathbf{X}\mathbf{X}^\top))^{-1}$  and

$\frac{1}{n} \mathbf{x}_{1:n}^\top \mathbf{y}_{1:n}$  is a consistent estimator of  $\mathbb{E}(\mathbf{X}\mathbf{Y}^\top).$

$\Rightarrow$  The least squares estimator  $(\mathbf{x}_{1:n}^\top \mathbf{x}_{1:n})^{-1} \mathbf{x}_{1:n}^\top \mathbf{y}_{1:n}$  is a consistent estimator of  $(\mathbb{E}(\mathbf{X}\mathbf{X}^\top))^{-1} \mathbb{E}(\mathbf{X}\mathbf{Y}^\top).$