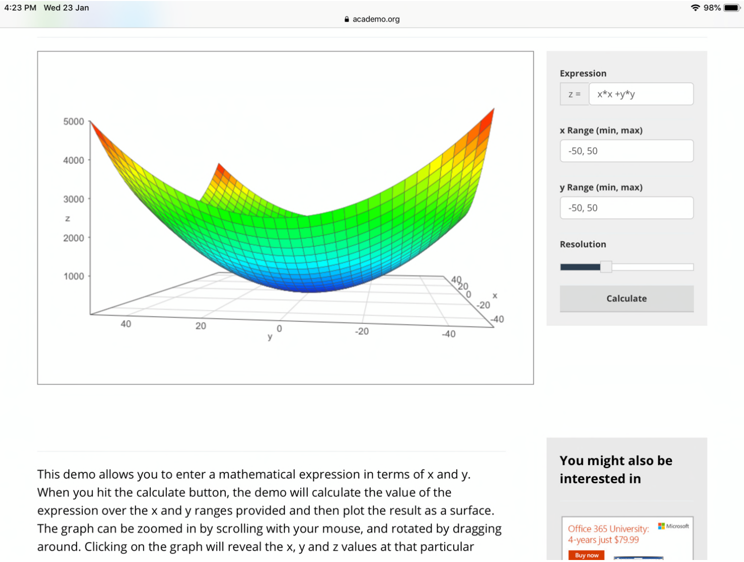
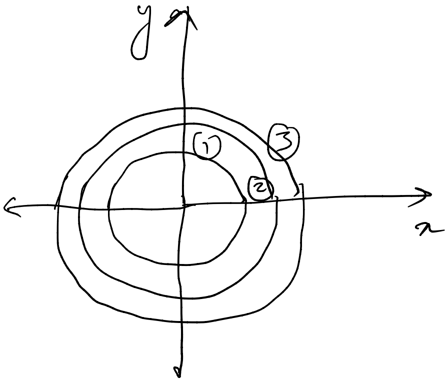
Mathematics for Machine learning – II

CONTOUR PLOT

$$f(x,y) = x^2 + y^2$$

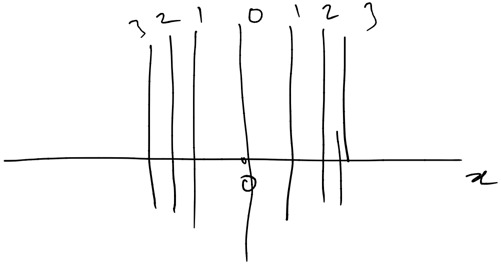
Plot $f(x,y) = K$ for varying K



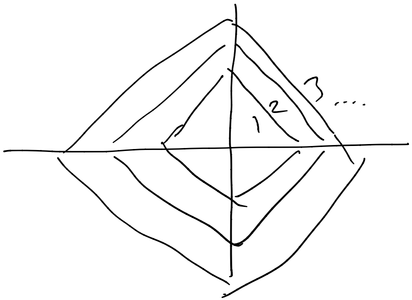


Q) Draw contour plot for $f(x) = x^2$

plot $x^2 = K$



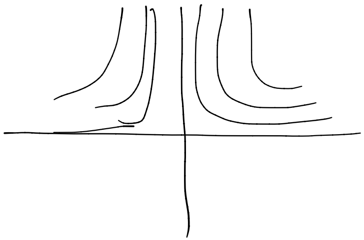
Q) Draw contour plot for for $f(x,y) = |x| + |y|$



$x, y \geq 0$

$\frac{x}{K} + \frac{y}{\cancel{K}} = 1$

Q) Draw contour for $f(x,y) = x^2 y$



$x^2 y = K$

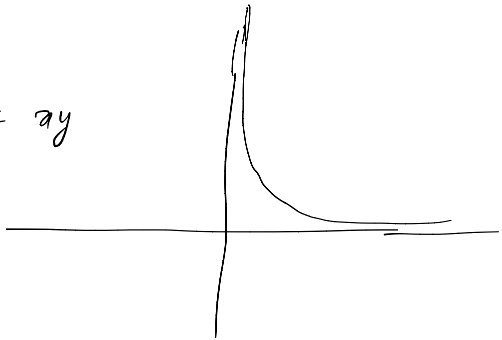
$x^2 = \frac{K}{y}$

or $y = \frac{K}{x^2}$

$K = 1$

$y = \frac{1}{x^2}$

Q) $f(x,y) = xy$



$xy = K \Rightarrow x = \frac{K}{y}$

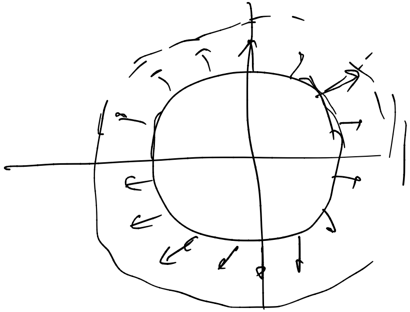
$K = 1$

$x = \frac{1}{y}$ $\quad$ $y = \frac{1}{x}$

Contour plots and (gradients)

↳ steepest change....

$f(x,y) = x^2 + y^2$



* All points on contour have same $f(x,y)$

* which dir^n to move to increase $f(x,y)$ more

⇒ ⊥ to given point on the curve

Gradient of : $\nabla f(x,y)$

$$= \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{bmatrix}$$

$$= \begin{bmatrix} 2x \\ 2y \end{bmatrix} = 2\begin{bmatrix} x \\ y \end{bmatrix}$$

Constrained optimisatⁿ

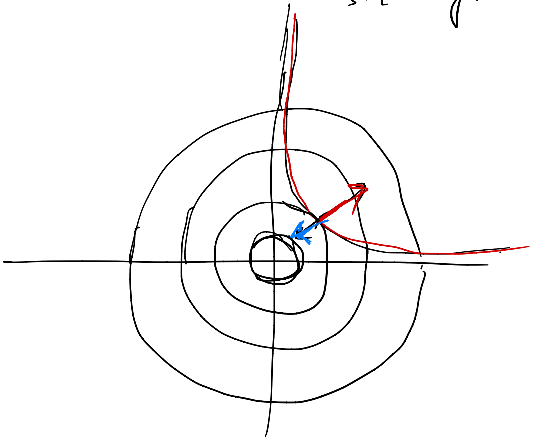
Extrema (max or minima) $f(x,y)$ x^2+y^2 s.t. $xy = K$

Q. Extrema is minima? or maxima?

∵ $x+y=1$
x large
y may be...

More generally, we may extrema $f(x...)$ s.t. $g(x...) = 0$

→ Gradient $f(x,y)$
→ Gradient of $g(x,y)$



$$\boxed{\nabla f(x,y) = \lambda (\nabla g(x,y))}$$

$$\nabla f(x,y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} = \lambda \nabla g(x,y) = \lambda \begin{bmatrix} y \\ x \end{bmatrix}$$

$$\Rightarrow \left.\begin{aligned} 2x &= \lambda y \\ 2y &= \lambda x \end{aligned}\right\} \quad \begin{aligned} &\text{3 variables} \quad \text{--- (1)} \\ &\text{2 eq}^{n} \quad \text{--- (2)} \end{aligned}$$

$$xy = K \quad \text{--- (3)}$$

(1) and (3)

$$x = \frac{\lambda y}{2} \ \& \ xy = K \Rightarrow \frac{\lambda y}{2} \times y = K \Rightarrow \lambda y^2 = 2K$$

(1) and (2)

$$4xy = \lambda^2 xy \Rightarrow \lambda = 2$$

$$2x = 2y \Rightarrow x = y \Rightarrow \boxed{\begin{aligned} x = y = \sqrt{K} \\ \lambda = 2 \end{aligned}}$$

Constrained optimisation

Extrema (max or minima) $\quad f(x,y) = x^2+y^2$

s.t. $\quad x+y=1$

Q. Extrema is minima? or maxima?

$\because$ $x+y=1$

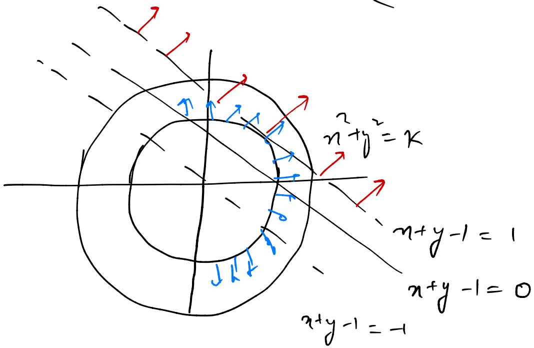
x incr.

y may dec.

More generally, we say $\left\{ \begin{array}{l} \text{Extrema } f(x \ldots) \\ \text{s.t. } g(x \ldots) = 0 \end{array} \right\}$

$\rightarrow$ Gradient $f(x,y)$

$\rightarrow$ Gradient of $g(x,y)$



All points on contour hand

$$\boxed{\nabla f(x,y) = \lambda (\nabla g(x,y))}$$

$$\nabla f(x,y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} = \lambda \nabla g(x,y) = \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \left.\begin{aligned} 2x &= \lambda \\ 2y &= \lambda \end{aligned}\right\} \text{2 eqn's \& 3 variables}$$

But we had

$x + y - 1 = 0$ ... 3rd eqn.

$\therefore$ $2x = \lambda$, $2y = \lambda$ $\Rightarrow$ $x = y = \frac{\lambda}{2}$

$x + y = 1 \Rightarrow \lambda = 1 \Rightarrow x = y = \frac{1}{2}$

λ - Lagrangian multiplier

Lagrangian $L(x, y, \lambda)$

$$\nabla L = 0$$

$$\frac{\partial L}{\partial x} = 0$$

$$\frac{\partial L}{\partial y} = 0$$

$$\frac{\partial L}{\partial \lambda} = 0$$

Q) Find extrema for $f(x,y) = x^2 y$

S.t. $g(x,y) = x^2+y^2-1=0$

$$L = x^2 y + \lambda (x^2+y^2-1)$$

$\frac{\nabla L}{\partial x} = 0 \Rightarrow 2xy + \lambda (2x) = 0$

$\frac{\partial L}{\partial y} \Rightarrow x^2 + \lambda (2y) = 0$

$\frac{\partial L}{\partial \lambda} \Rightarrow x^2+y^2-1=0$

$\Rightarrow x(y+\lambda) = 0$

$x^2 + 2\lambda y = 0$

$x^2+y^2 = 1$

Case I

$x = 0$

$\Rightarrow y^2 = 1 \Rightarrow y = \pm 1$

$\lambda = 0$

$f(x,y) = 0$

Case II

$x \neq 0 \Rightarrow y = -\lambda$

$\Rightarrow x^2 + 2\lambda(-\lambda) = 0 \Rightarrow x^2 = 2\lambda^2$

$x^2 + \lambda^2 = 1$

$\Rightarrow 3\lambda^2 = 1$

$\Rightarrow \lambda = \pm \frac{1}{\sqrt{3}}$

$$x^2 = 2\lambda^2 = \frac{2}{3}$$

$$y = \pm\sqrt{\frac{1}{3}}$$

$$\max x^2 y = \frac{2}{3}\sqrt{\frac{1}{3}}$$

KKT conditions

Minimize $f(x)$

s.t. $h_i(x) = 0 \quad \forall i = 1, \dots m$

$g_i(x) \leq 0 \quad \forall j = 1, \dots n$

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{n} \mu_i g_i(x)$$

Now, if $g_i(x^*) < 0$, then μ_i can be set to zero (constraint doesn't participate)

else $g_i(x^*) = 0$

$\therefore \boxed{\mu_i g_i(x^*) = 0}$

$\mu_i > 0$

Stationarity

$$\nabla_x f(x) + \sum_{i=1}^{m} \nabla_x \lambda_i h_i(x) + \sum_{i=1}^{n} \nabla_x \mu_i g_i(x) = 0$$

Equality

$$\nabla_\lambda f(x) + \sum_{i=1}^{m} \nabla_\lambda \lambda_i h_i(x) + \sum_{i=1}^{m} \nabla_\lambda \mu_i g_i(x) = 0$$

Inequality / Complementary slackness

$$\mu_i g_i(x) = 0$$

$$\mu_i \geq 0$$