LAGRANGE MULTIPLIERS

3/22/11

Suppose we want to minimize $f(x,y)$ subject to a constraint $g(x,y) = c$. The variables x y are not independent due to $g(x,y)$.

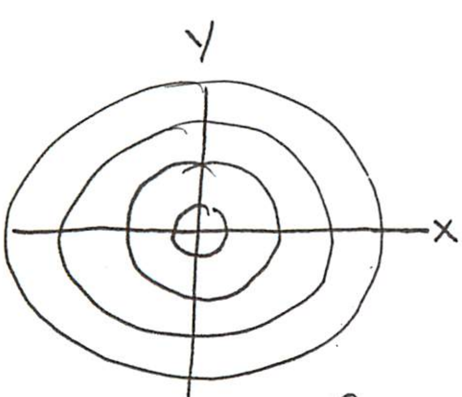
e.g. (squared distance)

$$\min_{x,y} \; x^2 + y^2 \quad \Leftarrow \quad f(x,y) = x^2 + y^2$$

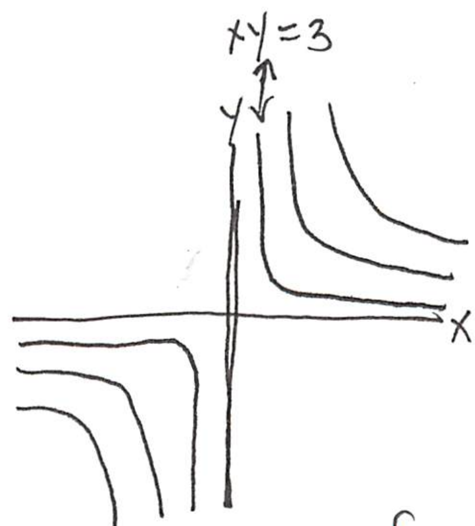
s.t.

$$xy = 3 \quad \Leftarrow \quad g(x,y) = xy = 3$$

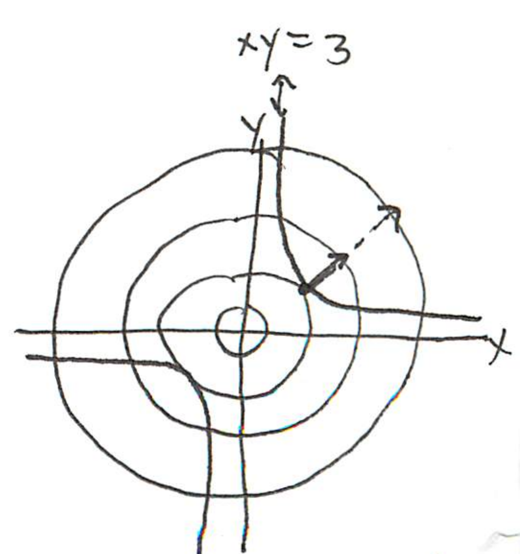
Recall the level curves of a function $h(x,y)$ are points $h(x,y) = c$ for various c.



Level curves of $f(x,y) = x^2 + y^2$



Level curves of $g(x,y) = xy$



Level curves of $f(x,y) = x^2 + y^2$ w/ ① Constraint curve $xy=3$

OBSERVE: At the minimum ~~of~~ level curve of $f(x,y)$ is tangent to hyperbola $g(x,y)=3$.

KEY IDEA: At stationary points the gradient of $f(x,y)$ is Parallel to the gradient of $g(x,y)$.

$$\nabla f \parallel \nabla g$$

So, $\nabla f = \lambda \cdot \nabla g$

We solve for all points where 1) gradients of $f(x,y)$ and $g(x,y)$ are parallel 2) Constraint is valid.

$$\begin{aligned} f_x &= \lambda g_x \\ f_y &= \lambda g_y \\ g(x,y) &= c \end{aligned} \iff \begin{aligned} 2x &= \lambda y \\ 2y &= \lambda x \\ xy &= 3 \end{aligned}$$

Three equations and three unknowns. We solve as a system of linear equations.

$$\begin{bmatrix} 2 & -\lambda \\ -\lambda & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

~~Three~~ Solutions: $(\sqrt{3}, \sqrt{3})$, $(-\sqrt{3}, -\sqrt{3})$

NOTE: No way to know whether a stationary point is a) minimum b) maximum c) saddle point. Have to plug-in and check.

e.g. (ML for Multinoulli)

$X_i \sim \text{Mult}(\theta)$ iid for $i = 1, \dots, N$

Log-Likelihood:

$$L(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{I}(x_i, k) \log \theta_k$$

Maximum Likelihood:

$$\max_{\theta} L(\theta)$$

s.t.

$$\sum_{k=1}^{K} \theta_k = 1 \quad , \quad \text{(normalization)}$$

Stationarity Conditions:

$$\nabla_\theta L(\theta) = \lambda \nabla_\theta \left(\sum_{k=1}^{K} \theta_k - 1 \right)$$

$$\sum_{k=1}^{K} \theta_k = 1$$

Partial derivatives are of the form,

$$\frac{\partial L}{\partial \theta_k} = \sum_{i=1}^{N} \cancel{\sum_{k=1}} \mathbb{I}(x_i, k) \frac{1}{\theta_k} = \frac{N_k}{\theta_k}$$

Stationary points satisfy,

$$\theta_k = \frac{N_k}{\lambda} \qquad (*)$$

$$\sum_k \theta_k = 1 \qquad (**)$$

Plugging $(*)$ into $(**)$ we have

$$\lambda = \sum_{k=1}^{K} N_k = N$$

and so,

$$\boxed{\theta_k = \frac{N_k}{N}}$$

Which is ML solution for multinoulli parameters.

<u>NOTE:</u> We did <u>not</u> impose nonnegativity constraint $\theta_k \geq 0$. In general, if we don't enforce a constraint $g(x)=c$ and the critical points are feasible, then they are optimal under constraint $g(x)=c$.

Inequality Constraints:

Consider inequality constraint $g(x) \geq c$.
This defines a set of feasible solutions called the "feasible region".

We must now consider solutions

1) On the interior $g(x) > c$. (inactive)
2) On the boundary $g(x) = c$. (active)

x_A - on boundary
x_B - interior

For solutions on boundary they are maximum only if gradients are opposing

$$\nabla f(x) = -\lambda \nabla g(x)$$

for $\lambda > 0$

And minimum if in the same direction,

$$\nabla f(x) = \lambda \nabla g(x) \qquad \text{for } \lambda \geq 0.$$

The stability conditions are known as KKT conditions. For minimizing,

$$\nabla f(x) = \lambda \nabla g(x)$$
$$g(x) \geq c$$
$$\lambda \geq 0$$

This is a generalization of Lagrange multipliers for inequality constraints.

<u>Note</u> <u>Differences</u>:

1) λ constrained to be positive
2) Direction of ∇f vs. ∇g important
3) Must consider active vs. inactive constraints.