Lagrange Multipliers Notes

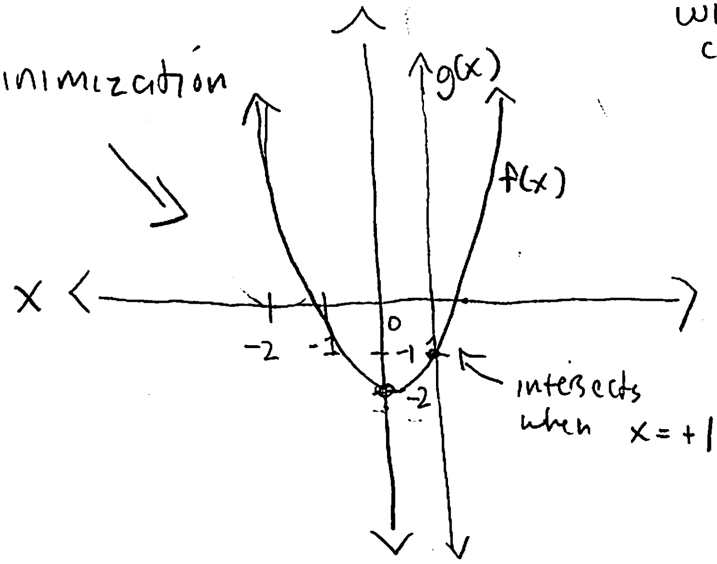
Modified from Jason Pacheco's notes from previous CS195F

minimize /maximize a function $f(x,y)$ subject to a constraint $g(x,y) = c$. The variables x, y are not independent due to $g(x,y)$.

One dimensional example:

Minimize $f(x) = x^2 - 2$
with constraint $g(x) = x - 1 = 0$

Simple minimization



Can be solved just by looking at the graph!

The optimal value for x given our constraint $g(x)$ is 1.

$$f(1) = 1^2 - 2 = -1$$
$$g(1) = 1 - 1 = 0 \checkmark$$ constraint is met.

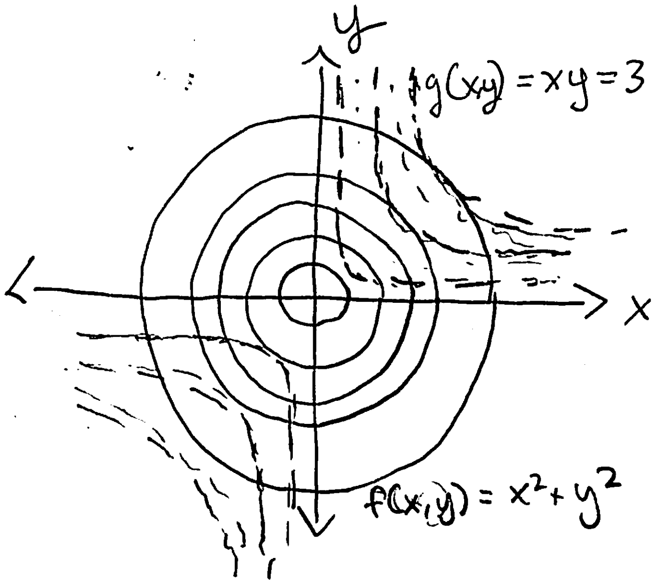
Plugging "1" back into the function is known as substitution (a really simple case). Solving for one equation and using that to solve for the other.

We now consider a 2-dimensional case:

Minimize $f(x,y) = x^2 + y^2$
subject to $g(x,y) = 3$

Visually what does this look like?

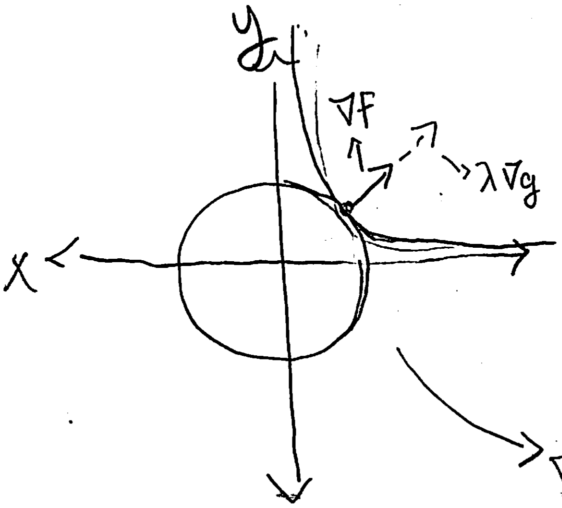
Level curves of $f(x) = x^2+y^2$



We want to find the smallest level set (i.e circle) that satisfies our constraint $g(x,y) = xy = 3$.

Key point

At stationary points (or where the derivative is 0) the gradient at $f(x,y)$ is parallel to the gradient of $g(x,y)$. Note that differentiable curves that touch one another but do not cross are tangential.



Key idea!!

$\nabla f \parallel \nabla g$

$\Downarrow$

$\nabla f = \lambda \nabla g$

$\Downarrow$

The problem becomes one where we solve for all points where

1) gradients of $f(x,y)$ and $g(x,y)$ are parallel

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

2) constraint is valid.

$g(x,y) = 0$ <= needs to be the case where the derivative w.r.t to λ sets the function $g(x)$ to be 0.

We can rewrite this now as a Lagrangian:

$$\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

and look for points where $\nabla \mathcal{L}(x,y,\lambda) = 0$

If the solution is on the boundary, the solution is a maximum only if gradients are opposing.

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

for $\lambda > 0$

minimum if in the same direction

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

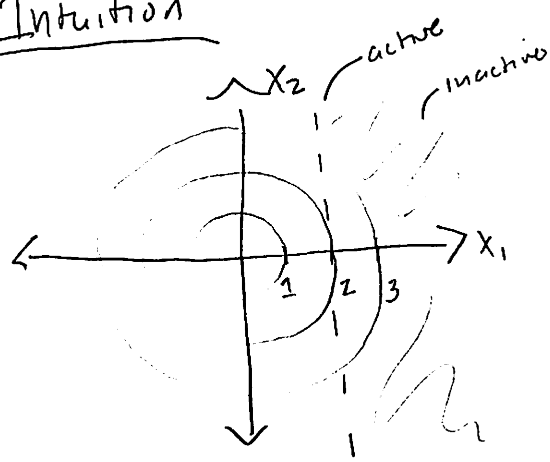
for $\lambda \geq 0$

These stability conditions are known as the KKT theorem (Karush - Kuhn - Tucker). Determines the form of these Lagrangian equations for an optimal solution.

Minimum	Maximum
$\nabla f(x) = \lambda \nabla g(x)$ ← equality constraint	$\nabla f(x) = \lambda \nabla g(x)$
$\nabla f(x) = \mu \nabla h(x)$ ← inequality constraint; ↑ another multiplie	$\nabla f(x) = \mu \nabla h(x)$ — the same but with different constraints
$\mathcal{L}(x,\lambda,\mu) = f(x) + \lambda \nabla g(x) + \mu \nabla h(x)$	$\mathcal{L}(x,\lambda,\mu) = f(x) + \lambda \nabla g(x) - \mu \nabla h(x)$
s.t $\mu \geq 0$	s.t $\mu \leq 0$
s.t. $\mu h(x) = 0$ (either $h(x)$ or μ has to be 0)	s.t $\mu h(x) = 0$

↑ also known as complementary slackness

Intuition



$$f(x) = x_1^2 + x_2^2$$

$$h(x) = x \geq 2 \Rightarrow x - 2 \geq 0$$

positive multipliers suggest that the direction of increasing $f(x)$ is in the same direction as increasing $h(x)$.

We can rewrite our previous problem in Lagrangian terms.

$$\mathcal{L}(x,y,\lambda) = F(x,y) + \lambda g(x,y) = x^2 + y^2 + \lambda(xy - 3)$$

Take partial derivatives w.r.t x, y, λ

$$\left.\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= 2x + \lambda y = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 2y + \lambda x = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= xy - 3 = 0\end{aligned}\right\}$$

3 equations w/ 3 unknowns

Solve this system of equations

Solve these equations to get: $(-\sqrt{3}, -\sqrt{3})$, $(+\sqrt{3}, +\sqrt{3})$

x_{min} y_{min} x_{min} y_{min}

$f(-\sqrt{3}, -\sqrt{3}) = 3 + 3 + (3-3) = 6$

For all values that satisfy the constraint

Inequality Constraints

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

The solutions of our Lagrangian now lie on

1) The interior $g(x) > c$ (inactive)

2) The boundary $g(x) = c$ (active)

Solving Multinomial MLE using Lagrange multipliers

Let θ_K be the parameter that defines the probability of categorical variable K.

Likelihood of multinomial

$$p(D|\theta) = \prod_{K=1}^{K} \theta_K^{n_K} \quad \text{wher} \quad n_K = \sum_{i=1}^{N} \mathbb{I}(y_i = k)$$

Constraint needs to be added to maintain a valid probability distribution.

$$\mathcal{L}(\theta, \lambda) = \sum_K n_K \log(\theta_K) + \lambda\left(1 - \sum_K \theta_K\right)$$

Take partial derivatives

$$\frac{\partial \mathcal{L}}{\partial \theta_K} = \frac{n_K}{\theta_K} - \lambda = 0 \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_K \theta_K = 0$$

Solve for θ_K

$$\theta_K = \frac{n_K}{\lambda}$$

$$1 - \sum_K \frac{n_K}{\lambda} = 0$$

$$\sum_K \frac{n_K}{\lambda} = 1 \Rightarrow \frac{1}{\lambda} \sum_k n_k = 1$$

$$\frac{1}{\lambda} N = 1 \Rightarrow \lambda = N$$

$$\hat{\theta}_K = \frac{n_K}{N}$$

Solving $L(x, y, \lambda) = x^2 + y^2 + \lambda(xy - 3)$

We know that

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda y = 0 \quad \Bigg| \quad \frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda x = 0 \quad \Bigg| \quad \frac{\partial \mathcal{L}}{\partial \lambda} = xy - 3 = 0$$

Solve for λ

$$\lambda y = -2x$$

$$\lambda = \frac{-2x}{y}$$

subs.

$$2y + \left(\frac{-2x}{y}\right)x = 0$$

$$2y - \frac{2x^2}{y} = 0$$

$$2y^2 = 2x^2$$

$$y^2 = x^2$$

Since $x = y$ we know that only

$$x = \pm\sqrt{3}$$

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