Lagrange multipliers

Problem Statement

Consider function $f(x)$

$$\underset{\text{Input}}{\mathbb{R}^D} \rightarrow \boxed{f} \rightarrow \underset{\text{Output}}{\mathbb{R}}$$

$$x \qquad\qquad\qquad f(x)$$

Goal: constrained optimization

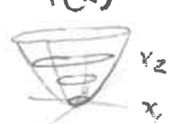
focus on equality constraints

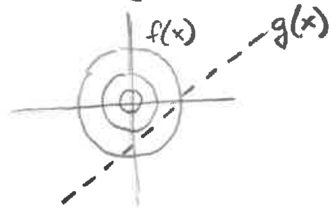
Find $x^* = \underset{x}{\operatorname{argmin}} f(x)$

subject to constraint $g(x) = 0$

see p. 7 for inequality

"bowl" function $f(x)$





Find the x on line defined by g that minimizes $f(x)$

Recipe

1) Define objective function f and constraints g

2) Define expanded objective function

introduced new "multiplier" variable λ

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

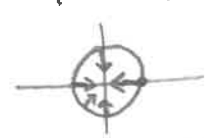
3) Set up system of eqs, where each partial deriv equals zero

$$\frac{\partial}{\partial x_1}\mathcal{L} = 0, \quad \frac{\partial}{\partial x_2}\mathcal{L} = 0, \quad \ldots \quad \frac{\partial}{\partial x_D}\mathcal{L} = 0, \quad \frac{\partial}{\partial \lambda}\mathcal{L} = 0$$

4) solve system of equations to find the optimal value of x

Writing Constraints as Functions $g(x)$

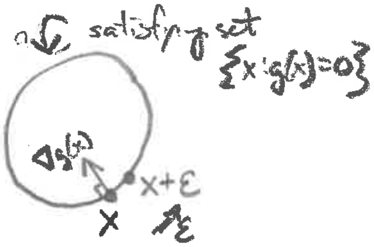
Here are some examples of turning an <u>equality</u> constraint into a function $g(x)$ that is <u>zero</u> when inputs satisfy constraint.

	Constraint	Function $g(x)$	Gradient $\nabla g(x)$
1D	Find x (a scalar) "such that" s.t. $x - 1 = 0$	$g(x) = x - 1$	
2D	Find $x = [x_1, x_2]$ s.t. $x_1 = x_2$	$g(x) = x_1 - x_2$ or $g(x) = x_2 - x_1$	
2D	Find $x = [x_1, x_2]$ s.t. x lies on line with slope 1 and intercept 3 i.e. $x_2 = 1 x_1 + 3$	$g(x) = x_2 - (x_1 + 3)$ 	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$
2D	Find $x = [x_1, x_2]$ s.t. x lies on unit circle $x_1^2 + x_2^2 = 1$	$g(x) = 1 - x_1^2 - x_2^2$ 	$\begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$
N-dim	Find $x = [x_1, x_2 \cdots x_D]$ s.t. $\sum_d x_d = 1$	$g(x) = 1 - \sum_d x_d$	

Intuition for why it works

Step 1: Realize that any x that satisfies constraints must have gradient perpendicular to constraint surface

Consider x, $x+\varepsilon$ that both satisfy constraints



(1) $$g(x) = 0$$
$$g(x+\varepsilon) = 0$$

Taylor thm says

(2) $$g(x+\varepsilon) \approx g(x) + \varepsilon^T \nabla g(x)$$

Alternatively, realize satisfying set is isocontour of function g, so by def. all points have perp. gradient

Plugging in the zeros from (1) gives

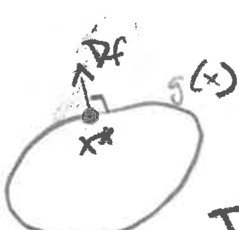
$$0 = \varepsilon^T \nabla g(x)$$

Dot product of zero implies vectors are perpendicular

so $\nabla g(x)$ is orthogonal to a step along constraint surface

Step 2: At optimal point x^*, both objective's gradient and constraint gradient are perpendicular to constraint surface and parallel to each other

Suppose ∇f was not orthogonal at x^*, then we could move along constraint surface & improve f, which means x^* is not optimal



both ∇f and ∇g are orthogonal to same surface

$\Downarrow$

$\nabla f, \nabla g$ are parallel $\qquad \nabla f + \lambda \nabla g = 0 \quad$ for some $\lambda \neq 0$

Intuitive Proof: why it works

(cont'd from prev page)

Step 3: Set up system of equations s.t. solution x^*, λ satisfies optimization problem

---

Need to guarantee

(1) $\nabla_x f + \lambda \nabla_x g = 0$ at x^*

(2) $g(x^*) = 0$

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

Partial wrt x satisfies (1)

Partial wrt λ satisfies (2)

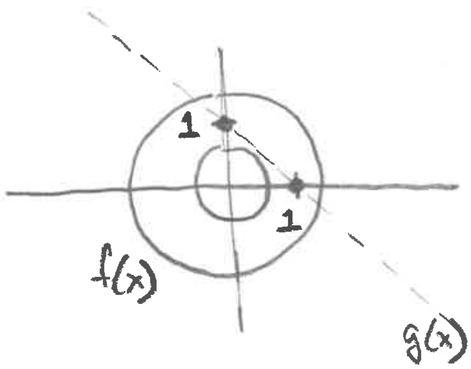
Example

$X = [x_1 \quad x_2]$

Step 1

$f(x) = 1 - x_1^2 - x_2^2$ $\qquad$ $g(x) = x_1 + x_2 - 1$

level sets are circles $\qquad$ just a line



Intuitively
solution is point on line closest to origin.

$x^* = [\tfrac{1}{2} \quad \tfrac{1}{2}]$

Via Lagrange multipliers

Step 2

$$\mathcal{L}(x, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = -2x_1 + \lambda \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = x_1 + x_2 - 1$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = -2x_2 + \lambda$$

Step 3 $\quad$ Set up system of eqs

$$(1) \quad 0 = -2x_1 + \lambda$$
$$(2) \quad 0 = -2x_2 + \lambda$$
$$(3) \quad 0 = x_1 + x_2 - 1$$

Step 4 $\quad$ Solve for x, λ

(1) says $\lambda = 2x_1$, so then

$$0 = -2x_2 + 2x_1 \qquad x_1 = x_2 \Rightarrow x_2 = \tfrac{1}{2}$$
$$+\ 0 = 2x_1 + \cancel{2x_2} - 2$$
$$0 = 4x_1 - 2 \Rightarrow x_1 = \tfrac{1}{2}$$

$$\boxed{\begin{aligned} x^* &= [\tfrac{1}{2}, \tfrac{1}{2}] \\ \lambda &= 1 \end{aligned}}$$

Example:

ML estimation for parameters of categorical r.v.

Model

K possible discrete outcomes for random variable x_i

Prob. of k^{th} outcome $= \theta_k$

So

$$p(x_i = k \mid \theta) = \theta_k$$

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

Problem

Observed data: $\{x_1, x_2, \cdots x_N\}$

where $x_i \in \{1, 2, \cdots K\}$

Find θ that maximizes likelihood

$$\log p(x \mid \theta) = \sum_{k=1}^{K} \sum_{i=1}^{N} \{x_i = k\} \log \theta_k$$

$N_k \triangleq \sum_{i=1}^{N} \{x_i = k\}$

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

Intuitively

$$\theta_k = \frac{N_k}{\sum_{\ell} N_{\ell}}$$

Solution

$\boxed{\text{Step 1}}$ $f(\theta) = \sum_{k=1}^{K} N_k \log \theta_k \quad g(\theta) = 1 - \sum_k \theta_k$

$\boxed{\text{Step 2}}$ $\mathcal{L}(\theta, \lambda) = \sum_k N_k \log \theta_k + \lambda(1 - \sum_k \theta_k)$

$\boxed{\text{Step 3}}$ $\frac{\partial \mathcal{L}}{\partial \theta_k}: \frac{N_k}{\theta_k} - \lambda = 0 \quad \frac{\partial}{\partial \lambda}[\mathcal{L}]: 1 - \sum_k \theta_k = 0$

$\boxed{\text{Step 4}}$ $\theta_k = \frac{N_k}{\lambda} \Rightarrow 0 = 1 - \sum_k \frac{N_k}{\lambda}$

$\frac{1}{\lambda} \sum_k N_k = 1 \Rightarrow \lambda = \sum_k N_k$

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

Generalizations / Extensions

Multiple constraints

Write: $g_1(x), g_2(x), \cdots g_N(x)$

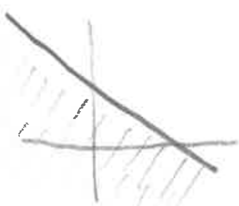
Set "expanded" lagrange objective

$$\mathcal{L}(x, \lambda) = f(x) + \lambda_1 g_1(x) + \cdots \lambda_N g_N(x)$$

Follow same recipe as before to find $x, \lambda_1, \cdots \lambda_N$

Inequality constraints

$x_2 \leq -x_1 + 3$



or $x_1^2 + x_2^2 \leq 3$

solutions are of two kinds

<u>active</u> (solution at boundary of feasible region)

<u>inactive</u> solution at <u>interior</u>

So to find max $f(x)$ s.t. $g(x) \geq 0$

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

subject to

$$g(x) \geq 0$$
$$\lambda > 0$$
$$\lambda g(x) = 0$$

KKT <u>conditions</u>

not as easy as <u>equality</u> constraints