PENALTY AND BARRIER FUNCTIONS

1. Penalty Method

Problem: $\min f$ under the constraint $g(x) \leq 0$.

Consider the constraint $g(x) \leq 0$. The idea of penalty is to have

$$P(x) = \begin{cases} 0 & g(x) \leq 0 \\ > 0 & g(x) > 0 \end{cases}$$

This can be achieved using the operation

$$\max(0, g(x))$$

which returns the maximum of the two values. We can make the penalty more regular by using

$$(\max\{g(x_1, x_2, \ldots, x_N), 0\})^2.$$

This is the quadratic penalty function.

In general

$$(\max\{g(x_1, x_2, \ldots, x_N), 0\})^p \quad p \geq 1$$

- $p = 1$ linear penalty function: this function may not be differentiable at points where $g(x) = 0$.
- $p = 2$. This is the most common penalty function.

Given a function $g^+(x_1, \ldots, x_N) = \max\{g(x_1, x_2, \ldots, x_N), 0\}$ with $g \in C^1$ then $\phi(x) = (\max\{g(x), 0\})^2$ is C^1 and

$$D\phi(x) = \begin{cases} 2g(x)Dg(x) & \text{if } g(x) > 0 \\ 0 & \text{if } g(x) \leq 0 \end{cases}$$

Hence

$$D\phi(x) = 2g^+(x)Dg(x).$$

Penalty method

Penalty method replaces a constrained optimization problem by an unconstrained problems whose solutions ideally converge to the solution of the original constrained problem. First we have converted the constraints into penalty functions, then we add all the penalty functions on to the original objective function and minimize from there: minimize

$$F_k(x) = f(x) + \frac{k}{2}(\max\{g(x), 0\})^2$$

We multiply the quadratic penalty function by $\frac{k}{2}$. The factor $k > 0$ controls how severe the penalty is for violating the constraint.

Solve the minimum problem under the constraint $g \leq 0$

$$\min f(x_1, x_2) = \|x\|^2 \qquad x = (x_1, x_2) \in \mathbb{R}^2$$
$$g(x) = x_1 + x_2 - 2 \leq 0$$

We consider

$$g^+(x_1, x_2) = \begin{cases} x_1 + x_2 - 2 & x_1 + x_2 - 2 > 0 \\ 0 & x_1 + x_2 \leq 2 \end{cases} \tag{1.1}$$

Introduce an artificial penalty for violating the constraint: we are trying to minimize f hence we add value when the constraint is violated.

$$F_k(x) = f(x) + \frac{k}{2}(g^+(x))^2 \quad , \quad k = 1, 2, \dots$$

$$F_k(x) = x_1^2 + x_2^2 + \frac{k}{2}\big(\max((x_1 + x_2 - 2), 0)\big)^2 \quad k = 1, 2, \dots$$

Making the gradient

$$\begin{cases} \frac{\partial F_k}{\partial x_1} = 2x_1 + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \\ \frac{\partial F_k}{\partial x_2} = 2x_2 + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \end{cases}$$

$$x_2 = x_1$$

$$x_1 = -k \max(x_1 - 1, 0) = \begin{cases} -k(x_1 - 1) & x_1 - 1 > 0 \\ 0 & x_1 - 1 \leq 0 \end{cases}$$

$$x_2 = -k \max(x_2 - 1, 0) \quad k = 1, 2, \dots$$

- Assume $x_1 - 1 > 0$, $x_2 - 1 > 0$ then $(1 + k)x_1 = k$ $x_1 = x_2 = \frac{k}{1+k}$ (not admissible since we assume $x_1 - 1 > 0$, $x_2 - 1 > 0$)
- Assume $x_1 - 1 \leq 0$, $x_2 - 1 \leq 0$ then $x_1 = x_2 = 0$

The solution is

$$x_1 = x_2 = 0$$

Solve the minimum problem under the constraint $g \leq 0$

$$\min f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$$
$$g(x) = x_1 + x_2 - 2 \leq 0$$

$$F_k(x) = f(x) + \frac{k}{2}(g^+(x))^2$$

$$F_k(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + \frac{k}{2}\big(\max((x_1 + x_2 - 2), 0)\big)^2 k = 1, 2, \dots$$

$$\begin{cases} \frac{\partial F_k}{\partial x_1} = 2(x_1 - 1) + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \\ \frac{\partial F_k}{\partial x_2} = 2(x_2 - 1) + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \end{cases}$$

$$x_2 = x_1$$

$$x_1 - 1 = -k\max(x_1 - 1, 0) = \begin{cases} -k(x_1 - 1) & x_1 - 1 > 0 \\ 0 & x_1 - 1 \le 0 \end{cases}$$

$$x_2 - 1 = -k\max(x_2 - 1, 0) \;\; k = 1, 2, \dots$$

- Assume $x_1 - 1 > 0$, $x_2 - 1 > 0$ then $x_1 = x_2 = 1$ (not possible since we assume $x_1 - 1 > 0$, $x_2 - 1 > 0$)
- Assume $x_1 - 1 \le 0$, $x_2 - 1 \le 0$ then $x_1 = x_2 = 1$.

The solution is

$$x_1 = x_2 = 1$$

.

Solve the minimum problem under the constraint $g \le 0$

$$\min f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2$$

$$g(x) = x_1 + x_2 - 2 \le 0$$

$$F_k(x) = f(x) + \frac{k}{2}(g^+(x))^2$$

$$F_k(x) = (x_1 - 1)^2 + (x_2 - 2)^2 + \frac{k}{2}\big(\max((x_1 + x_2 - 2), 0)\big)^2$$

$$\begin{cases} \frac{\partial F_k}{\partial x_1} = 2(x_1 - 1) + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \\ \frac{\partial F_k}{\partial x_2} = 2(x_2 - 2) + k\big(\max((x_1 + x_2 - 2), 0)\big) = 0 \end{cases}$$

$$x_2 - 2 = x_1 - 1$$

$$x_1 - 1 = -\frac{k}{2}\max(2x_1 - 1, 0)$$

$$x_2 - 2 = -\frac{k}{2}\max(2x_2 - 3, 0)$$

$x_1 - 1 + \frac{k}{2}(2x_1 - 1) = 0 \qquad (1 + k)x_1 = 1 + \frac{k}{2}$

$$x_1 = \frac{1 + \frac{k}{2}}{1 + k} \qquad x_2 = \frac{3\frac{k}{2} + 2}{k + 1}$$

$$k \to +\infty$$

$$x_1 = \frac{1}{2} \;\; x_2 = \frac{3}{2}$$

More generally, $f : \mathbb{R}^N \to \mathbb{R}$ penalty method for $\min_K f$ with $K : g_i(x) \le 0, \;\; i = 1, \dots M$ is

Set

$$P(x) = \sum_{i=1,\dots,M} \max\{0, g_i(x)\}^2$$

and minimize

$$\min[f(x) + \frac{k}{2}P(x) \;\; x \in \mathbb{R}^n \;\; k \in \mathbb{N}]$$

Go back to Lagrange multiplier method. The problem is the following Given $f : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R}^N \to \mathbb{R}^P$, find

$$\min\{f(x) : x \in \mathbb{R}^N \text{s.t.} h_i(x) = 0, \, i = 1, \ldots, P\} \tag{1.2}$$

Karush-Kuhn-Tucker conditions

$$\begin{aligned} &\min f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2 \\ &g(x) = x_1 + x_2 - 2 \leq 0 \end{aligned}$$

- Lagrangian
$$\mathcal{L}(x_1, x_2, \lambda) = (x_1 - 1)^2 + (x_2 - 2)^2 + \lambda(x_1 + x_2 - 2)$$
- Stationary condition
$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial}{\partial x_1}((x_1 - 1)^2 + (x_2 - 2)^2) + \lambda \frac{\partial}{\partial x_1}(x_1 + x_2 - 2) = 0$$
$$\frac{\partial \mathcal{L}}{\partial x_2} = \frac{\partial}{\partial x_2}((x_1 - 1)^2 + (x_2 - 2)^2) + \lambda \frac{\partial}{\partial x_2}(x_1 + x_2 - 2) = 0$$
- Admissibility (feasible) condition
$$x_1 + x_2 - 2 \leq 0$$
- Multiplier sign: non negativity of the multiplier
$$\lambda \geq 0$$
- Complementary slackness condition
$$\lambda(x_1 + x_2 - 2) = 0.$$

Find the solution. By the complementary slackness condition

$$\lambda(x_1 + x_2 - 2) = 0,$$

we have that $\lambda = 0$ or $x_1 + x_2 - 2 = 0$.

If $\lambda = 0$ then $\mathcal{L}(x_1, x_2, 0) = (x_1 - 1)^2 + (x_2 - 2)^2$, and

$$D\mathcal{L}(x_1, x_2, 0) = (2(x_1 - 1), 2(x_2 - 2)),$$

whose stationary point is $(1, 2)$. This is not an admissible point.

Let $x_1 + x_2 - 2 = 0$ then $x_2 = 2 - x_1$,

$$D_{x_1}\mathcal{L} = 2(x_1 - 1) + \lambda = 0$$

$$D_{x_2}\mathcal{L} = 2(x_2 - 2) + \lambda = 0,$$

then $x_2 = 2 - x_1$ and $x_1 - 1 = x_2 - 2$

$$x_1 = \frac{1}{2}, \quad x_2 = \frac{3}{2} \quad \lambda = 1$$

2. Barrier functions

In a constrained optimization a barrier function is a continuous function whose value on a point increases to infinity as the point approaches the boundary of the feasible region of an optimization problem. They are used to replace inequality constraints by a penalizing term in the objective function that is easier to handle.

Assumption: The set of strictly feasible points, $\{x : g_i(x) < 0, i = 1, ...m\}$ is nonempty.

$$\phi(x) = \sum_{i=1}^{M} \log(-g_i(x))$$

$$\nabla\phi(x) = \sum_{i=1}^{M} \frac{1}{g_i(x)} \nabla(g_i(x))$$

We consider

$$\min f(x) + \sum_{i=1}^{M} I_{g_i(x)\leq 0}(x)$$

$$I_{g_i(x)} = \begin{cases} +\infty & g_i(x) > 0 \\ 0 & g_i(x) \leq 0 \end{cases}$$

and the approximation by adding the log barrier function

$$F_\theta(x) = f(x) - \frac{1}{\theta} \sum_{i=1}^{M} \log(-g_i(x))$$

with θ a positive large number.

The idea in a barrier method is to avoid that points approach the boundary of the feasible region.

Next, we consider the minimization problem

$$\min[f(x) - \frac{1}{\theta} \sum_{i=1}^{M} \log(-g_i(x))],$$

$$g_i(x) < 0, \;\; i = 1, \dots M$$

whose stationary condition is

$$\theta\nabla f(x) - \sum_{i=1}^{M} \frac{1}{g_i(x)} \nabla(g_i(x)) = 0,$$

with condition

$$g_i(x) < 0, \;\; i = 1, \dots M$$

$c \in \mathbb{R}$ different from 0. We consider the minimization problem

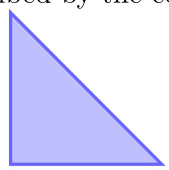
$$\min_K (cx + cy),$$

$x+y\le 1$, $x\ge 0$, $y\ge 0$.

We have $M=3$

$$g_1(x,y)=x+y-1\le 0$$
$$g_2(x,y)=-x\le 0$$
$$g_3(x,y)=-y\le 0$$

The domain K is described by the constraints $x+y\le 1$, $x\ge 0$, $y\ge 0$.

This is the feasible set. 

$$f(x,y)=cx+cy$$

We have $f(0,0)=0\quad f(0,1)=c\quad f(1,0)=c\quad f(x,y)=c$ if $x+y=1$.

If $c>0$ $f(0,0)=0$.

If $c<0$ $f(x,y)=c$ with $x+y=1$.

$c\in\mathbb{R}^n$.

$$\min[c^Tx-\frac{1}{\theta}\sum_{i=1}^{M}\log(-g_i(x))],$$

with g_i linear functions.

Fix $c\in\mathbb{R}$. We consider the minimization problem

$$\min_K(cx+cy),$$

and its approximation, $\theta>0$

$$\min[(cx+cy)-\frac{1}{\theta}(\log(-x-y+1)+\log(x)+\log(y)),$$

$x+y<1$, $x>0$, $y>0$.

$$F_\theta(x,y)=(cx+cy)-\frac{1}{\theta}(\log(-x-y+1)+\log(x)+\log(y))$$

Discuss the approximate problem.

$$F_\theta(x,y)=(cx+cy)-\frac{1}{\theta}(\log(-x-y+1)+\log(x)+\log(y))$$

Making the gradient

$$\theta c-\frac{1}{x+y-1}-\frac{1}{x}=0$$
$$\theta c-\frac{1}{x+y-1}-\frac{1}{y}=0.$$

$$\theta cx(x+y-1)-x-x-y+1=0$$
$$\theta cy(x+y-1)-y-x-y+1=0$$

Hence

$$\theta cx^2-(\theta c(1-y)+2)x+1-y=0,$$

$$\theta c y^2 - (\theta c(1-x)+2)y + 1 - x = 0.$$

Fix

$$\boxed{\theta c = t}.$$

Recall that θ is a positive large number

$$x^2 - ((1-y)+\frac{2}{t})x + \frac{1-y}{t} = 0,$$
$$y^2 - ((1-x)+\frac{2}{t})y + \frac{1-x}{t} = 0.$$

First we consider

$$x^2 - ((1-y)+\frac{2}{t})x + \frac{1-y}{t} = 0,$$
$$\Delta = ((1-y)+\frac{2}{t})^2 - 4\frac{1-y}{t} = (1-y)^2 + \frac{4}{t^2}$$
$$\sqrt{\Delta} = \sqrt{(1-y)^2 + \frac{4}{t^2}} = |1-y|\sqrt{1+\frac{4}{t^2(1-y)^2}}$$

For x small

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x$$
$$\sqrt{1+\frac{4}{t^2(1-y)^2}} \approx 1 + \frac{2}{t^2(1-y)^2}$$
$$x_{1,2} \approx \frac{1}{2}[(1-y)+\frac{2}{t} \pm (1-y)]$$
$$x_{1,2} \approx \begin{cases} (1-y)+\frac{1}{t} \\ \frac{1}{t} \end{cases}$$

Finally we get

$$\begin{cases} x+y \approx 1 + \frac{1}{\theta c} & c < 0 \quad \theta \text{ large.} \\ x = y \approx \frac{1}{\theta c} & c > 0 \quad \theta \text{ large.} \end{cases}$$