Non Linear programming

$$\min f(x) \qquad (P)$$
$$g_i(x) \le 0 \qquad i = 1 \dots r$$
$$h_j(x) = 0 \qquad j = 1 \dots m$$

As usual, define $F(P)$ as the set of feasible points (where all the constraints in P are satisfied).

<u>Thm</u> If $x^* \in F(P)$ is a local minimizer, there exist (Lagrange) multipliers $(\lambda, \mu) = (\lambda_0, \lambda_1, \dots, \lambda_r, \mu_1 \dots \mu_m)$ such that

$$\lambda_i \ge 0, \quad i = 0 \dots r; \quad (\lambda, \mu) \ne (0,0); \quad \text{and}$$
$$\lambda_0 \nabla f + \sum_j \lambda_j \nabla g_j + \sum_k \mu_k \nabla h_k = 0 \quad \text{at } x^*.$$

Moreover, $\lambda_j g_j(x^*) = 0$ for $j = 1 \dots r$.

<u>Proof</u> Consider convex cones:

$$K = \{ v : (\nabla f(x^*), v) < 0, \ (\nabla g_j(x^*), v) < 0, \ j \in I(x^*) \} \text{ and}$$
$$L = \{ v : (\nabla h_j(x^*), v) = 0, \ j = 1 \dots m \}$$

(NB: K is open; L is linear subspace). Here $I(x^*)$ are the indices of <u>active</u> constraints, i.e. $j \in I(x^*)$ iff $g_j(x^*) = 0$.

First, notice that the implications of the Theorem are satisfied if $\nabla h_j(x^*)$, $j = 1 \dots m$ are linearly dependent (just choose $\lambda = 0$; μ such that $\sum \mu_j \nabla h_j(x^*) = 0$, ✓).

If $\nabla h_j(x^*)$ are linearly independent, implicit function theorem conditions are satisfied, and (after, perhaps, reordering indices) we can find m functions ...

$$X_{n-m+1}(x_1 \dots x_{n-m}), X_{n-m+2}, \dots, X_n(x_1 \dots x_{n-m})$$

such that the set $\{h_j = 0,\ j = 1 \dots m\}$ is given by

$$x_k = X_k(x_1 \dots x_{n-m}),\ k = n-m+1, \dots, n, \text{ locally}$$

(i.e. $h_j(x_1, \dots, x_{n-m}, X_{n-m+1}(x_1, \dots, x_{n-m}), \dots, X_n(x_1 \dots x_{n-m})) \equiv 0$).

Now, assume the cones K and L have a nontrivial intersection, $v \in K \cap L$, $v \neq 0$. Consider the curve

$$x_k(t) = x_k^* + t v_k, \quad k = 1 \dots n-m;$$
$$x_k(t) = X_k(\tilde{x}^* + t\tilde{v}),\ k = n-m+1, \dots, n.$$

(here $\tilde{u} = (u_1, \dots, u_{n-m})$ for $u \in \mathbb{R}^n$).

It is immediate to check that thus defined curve

$$x(t) \in \mathbb{R}^n,\ x(0) = x^* \text{ lies in } \mathcal{F}(P) \text{ for}$$

small $t > 0$ (as all active inequalities have decreasing derivative at $t=0$, and all equalities h_j are satisfied by construction). Hence, if $K \cap L \neq \emptyset$, x^* is <u>not</u> a minimizer.

Hence, as x^* is, by assumption,

$$K \cap L = \emptyset, \text{ and therefore, there}$$

exists a vector $\bar{u} \in K^*$: $(\bar{u}, \bar{v}) \geq 0$ for all $\bar{v} \in L$.

However, any such vector would be a positive combination of vectors $\nabla f(x^*)$, $\nabla g_j(x^*)$ $j \in I(x^*)$ (by Dubovitsky-Milyutin thm) and also, as $(\bar{u}, \bar{v}) = 0$ for all $\bar{v} \in L$, $\bar{u}$ is a <u>linear</u> combination of $\nabla h_j(x^*)$. Hence

$$0 = \bar{u} - \bar{u} = \lambda_0 \nabla f + \sum \lambda_j \nabla g_j - \sum \mu_j \nabla h_j \qquad \square$$

<u>Definition</u> $L : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^m \to \mathbb{R}$: Lagrange function.

$$L(x, \lambda, \mu) = \lambda_0 f(x) + \sum \lambda_j g_j(x) + \sum \mu_j h_j(x)$$

<u>Examples</u> • Find vector closest to $b \in \mathbb{R}^n$, subject to bound on its ℓ_1 norm: $\frac{1}{2}|x-b|^2 \to \min$, $|x|_1 = \sum |x_k| \le c$

Making it smooth: for any binary vector $\varepsilon \in \{-1,1\}^n$,

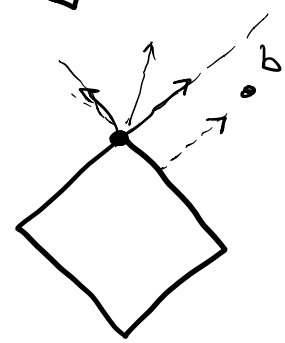
$$(\varepsilon, x) \le c.$$

Lagrange multipliers: λ_ε, $\varepsilon \in I_2^n$

$$L(x,\lambda) = \frac{1}{2}|x-b|^2 + \sum \lambda_\varepsilon \left[(\varepsilon, x) - c\right]$$

Necessary conditions of optimality:

$$\begin{cases} \text{for all } \varepsilon: \ (\varepsilon, x) = c, \ \lambda_\varepsilon = 0; \\ \lambda_0(x-b) + \sum \lambda_\varepsilon \varepsilon = 0. \end{cases}$$



• $y \to \min$ subj to

$$y - 2x^2 \le 0$$
$$x^2 - y \le 0$$

Lagrange function: $\lambda_0 y + \lambda_1(y - 2x^2) + \lambda_2(x^2 - y)$;

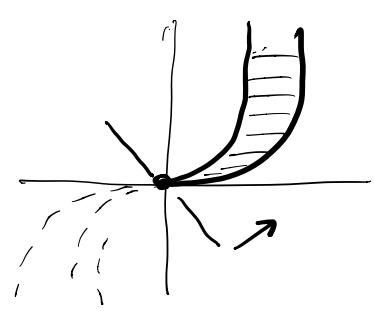
Necessary conditions for minimizer:

$$\lambda_0(0,1) + \lambda_1(-4x, 1) + \lambda_2(2x, -1) = 0, \ \lambda_j \ge 0 \ldots$$

$$\begin{bmatrix} \lambda_1 = 2\lambda_2 \\ \text{or } x = 0 \end{bmatrix}, \ \begin{bmatrix} \lambda_0 + \lambda_1 = \lambda_2 \end{bmatrix} \implies x = 0; \ y = 0.$$

• $x + y \to \min$ subj to

$$y - 2x^3 \le 0$$
$$x^3 - y \le 0$$



$$\lambda_0(1,1) + \lambda_1(-6x^2, 1) + \lambda_2(3x^2, -1) = 0$$

Useful view at Lagrange multiplier: penalty function.

Define $g^+ = \max(g, 0)$

Consider $F = f(x) + \frac{k}{2}\left[\sum g_j^+(x)^2 + \sum h_j(x)^2\right] + \frac{\gamma}{2}(x - x_*)^2$

Then x^* is a strict global minimizer in some ball $B_\rho(x^*)$: the value at x^* is $f(x^*)$, and if x^* is a local minimizer, then there is a global minimizer in $B_\rho(x^*) \cap \mathcal{F}(P)$, as then $F(x) > f(x^*) = F(x^*)$ for $x \in \partial B_\rho(x^*)$.

Let x_k be an unconstrained minimizer for F in $\mathring{B}_\rho(x^*)$. As $\inf_{x \in B_\rho} f \leq F(x_k) \leq f(x^*)$, the terms $g_j(x_k)$, $h_j(x_k) \to 0$. Take $\bar{x}$ to be a limit of x_k's along a subsequence. Then $\bar{x} \in B_\rho \cap \mathcal{F}(P)$, and $f(\bar{x}) \geq f(x^*)$. So, also $|\bar{x} - x^*|^2 = 0$, i.e. there is a sequence $x_k \to x^*$.

At x_k,

$$0 = \nabla F(x_k) = \nabla f + \sum k g_j^+ \cdot \nabla g_j(x_k) + \sum k h_j \nabla h_j(x_k) + (x_k - x^*).$$

We can rescal these coefficients so that they sit on the sphere, so that

$$0 = \lambda_{0,k} \nabla f + \sum \lambda_{j,k} \nabla g_j(x_k) + \sum \mu_{j,k} \nabla h_j(x_k) + \gamma_k (x_k - x^*)$$

Taking limits, we get the desired. $\square$