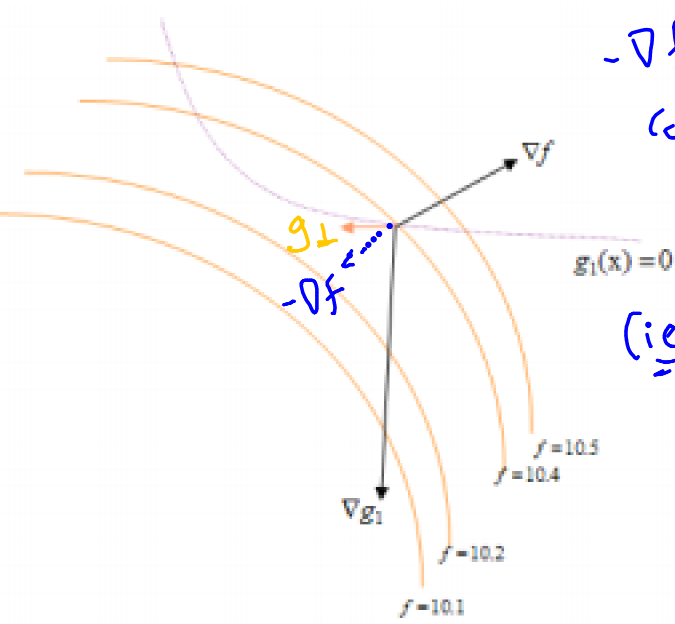
NECESSARY CONDITIONS FOR CONSTRAINED OPTIMALITY (pages 284-287 of...
http://www.cse.iitb.ac.in/~cs709/notes/BasicsOfConvexOptimization.pdf)

min $f(x)$
s.t. $g_1(x) = 0$



$-\nabla f$ has a small component $\perp$ to $\nabla g_1(x)$ ($g_\perp$) (ie $g_1(x)=0$ remains)

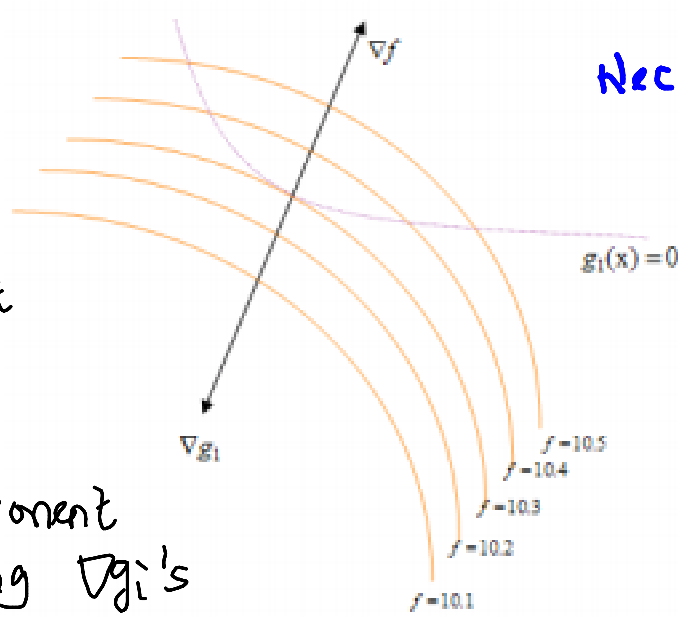
min $f(x)$
s.t. $g_i(x) = 0 \quad i = 1..m$

Figure 4.39: At any non-optimal and non-saddle point of the equality constrained problem, the gradient of the constraint will not be parallel to that of the function.

Necessary condition for multiple constraints

$\nabla f + \sum_i \lambda_i \nabla g_i = 0$ at x pt of min/max

ie. ∇f has no component $\perp$ to subspace containing ∇g_i's



Necessary condition

$\nabla f + \lambda \nabla g = 0$

at x pt of minimum/max

Figure 4.40: At the equality constrained optimum, the gradient of the constraint must be parallel to that of the function.

In the case of

$\min f(x)$

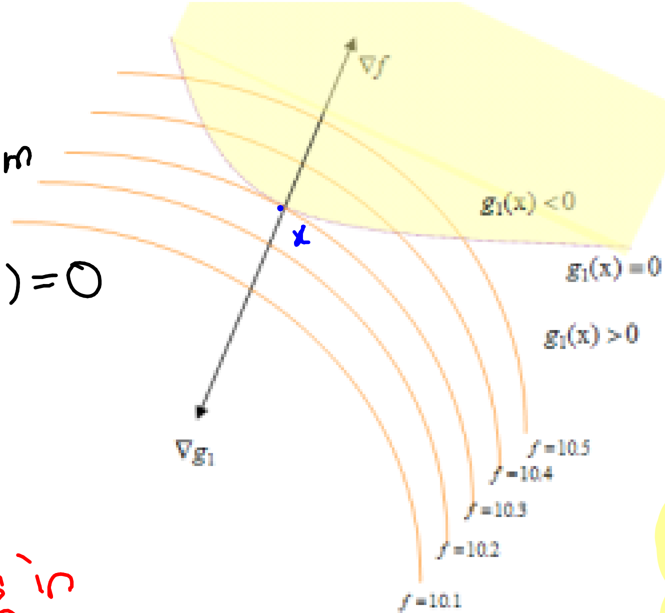
s.t. $g_i(x) \leq 0 \quad i = 1..m$

$\nabla f(x) + \sum_i \lambda_i \nabla g_i(x) = 0$

$\lambda_i g_i(x) = 0$

$\lambda_i \geq 0$

(See Recession cones in Bertsekas)



We do not mind reducing f & g_1 simultaneously

$\therefore -\nabla f(x)$ should not have component $\perp$ to ∇g_1 & also should not have component along $-\nabla g_1(x)$ (since $g_1(x) < 0$ is o.k.)

New additional condition

Figure 4.41: At the inequality constrained optimum, the gradient of the constraint must be parallel to that of the function.

$\nabla f(x) + \lambda \nabla g_1(x) = 0 \quad \lambda_1 \geq 0$

& $\lambda g_1(x) = 0$

This discussion of necessary conditions for "local" optimality holds also for non-convex, differentiable f & g_i's

We have invoked:

$$\left.\begin{aligned} &\min f(x) \\ &\text{s.t. } g_i(x) \leq 0 \quad i = 1..m \\ &h_j(x) = 0 \quad j = 1..p \end{aligned}\right\} (*)$$

$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$

Necessary condition for optimality of $(*)$ at x^* is

$$(**) \left\{\begin{aligned} &\nabla_x L(x^*, \lambda, \nu) = 0 \text{ for some } \lambda_i \geq 0 \text{ \& } \nu_j\text{'s (no constraints on } \nu_j\text{'s)} \\ &\text{\&} \quad \lambda_i g_i(x^*) = 0 \\ &g_i(x^*) \leq 0 \quad h_j(x^*) = 0 \end{aligned}\right.$$

Q1

$$\left\{\begin{aligned} &L^*(\lambda, \nu) = \min_x L(x, \lambda, \nu) \\ &\lambda^*, \nu^* = \underset{\lambda \geq 0, \nu}{\operatorname{argmax}} L^*(\lambda, \nu) \end{aligned}\right.$$

Could we say (when) that λ^*, ν^* that maximize the dual fn $L^d(\lambda, \nu)$ are precisely the λ's & ν's that satisfy the necessary conditions $*\Delta$?

Q2 If any λ^*, ν^*, x^* satisfy $*\Delta$, should the duality gap be 0 ?

Q3 Do answers to Q1 & Q2 require convexity of f & g_i's & affiness of h_j's

Boyd uses slightly different notations in the following slides

(WE)

$$\min_x \; f(x)$$
$$\text{s.t. } g_i(x) \le 0 \quad i = 1..m$$
$$h_j(x) = 0 \quad j = 1..p$$

(BOYD)

$$\min_x \; f_0(x)$$
$$\text{s.t. } f_i(x) \le 0 \quad i = 1..m$$
$$h_j(x) = 0 \quad j = 1..p$$

Lagrange function

$$L(x,\lambda,\nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Dual function

$$L^*(\lambda,\nu) = \min_x L(x,\lambda,\nu)$$

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu)$$