Maximize the margin

p. 111: we are looking for is

$$\tilde{\underline{w}}, \tilde{b} = \arg\max_{\underline{w}, b} \left\{ \min_{\ell} \left[\frac{t_\ell (\underline{w}^T \underline{y}_\ell + b)}{\|\underline{w}\|} \right] \right\}$$

$$= \arg\max_{\underline{w}, b} \left\{ \frac{1}{\|\underline{w}\|} \min_{\ell} \left[t_\ell (\underline{w}^T \underline{y}_\ell + b) \right] \right\}$$

Highly complex optimization problem!

Observation: problem can be simplified by considering that after rescaling $\underline{w} \to \kappa \underline{w}$ and $b \to \kappa b$, the distance from any point $\underline{y}_\ell$ to the decision surface does not change!

$\Rightarrow$ gives a degree of freedom to set

$t_\ell (\underline{w}^T \underline{y}_\ell + b) = 1$

for the points closest to the decision surface (ACTIVE CONSTRAINTS)

and $t_\ell (\underline{w}^T \underline{y}_\ell + b) > 1$

for other points (INACTIVE CONSTRAINTS)

$\Rightarrow$ overall : $\boxed{t_\ell (\underline{w}^T \underline{y}_\ell + b) \geq 1 \text{ for all } \ell}$

CANONICAL REPRESENTATION OF DECISION HYPERPLANE!

$\Rightarrow$ As a result, it is required (for maximizing the margin AND keeping the constraints) then we would maximize $\frac{1}{\|\underline{w}\|}$ subject to

(s.t.) $t_\ell (\underline{w}^T \underline{y}_\ell + b) \geq 1$ for $\ell = 1, \ldots, L$.

$\Rightarrow$ equivalent to

$$\underset{\underline{w}, b}{\text{maximize}} \; \frac{\|\underline{w}\|^2}{2} \quad \text{s.t. constraints}$$

How do we maximize a function s.t. constraints (inequality constraints, but can also include equality constraints (cf. Bachelor classes in analysis)) ?

$\Rightarrow$ Cf. App. E (Bishop's book) on Lagrange multipliers)

First consider the case of equality constraints for two-dimensional parameter vectors $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, where we are trying to maximize a function $f(\underline{x})$ s.t. $g(\underline{x}) = 0$

Define Lagrangian function

$$\boxed{L(\underline{x}, \underset{\text{real}}{\lambda}) \equiv f(\underline{x}) + \lambda g(\underline{x})}$$

$\Rightarrow$ maximize $L(\underline{x}, \lambda)$ w.r.t. both $\underline{x}$ and λ:

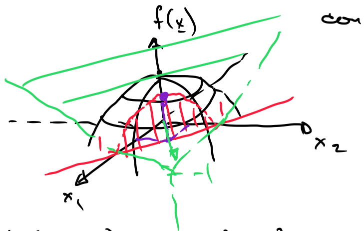
$$\boxed{\nabla_{\underline{x}} L = \underline{0} \quad \wedge \quad \frac{\partial L(\underline{x}, \lambda)}{\partial \lambda} = g(\underline{x}) = 0}$$

Example: $f(\underline{x}) = 1 - \|\underline{x}\|^2 = 1 - x_1^2 - x_2^2$

constraint: $g(\underline{x}) = x_1 + x_2 - 1 = 0$

$x_2 = -x_1 + 1$

$g(\underline{x}) = x_1 + x_2 - 1$



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

$\nabla_{\underline{x}} L = \underline{0}$: $\left.\begin{aligned} -2x_1 + \lambda &= 0 \\ -2x_2 + \lambda &= 0 \end{aligned}\right\} \rightarrow x_1 = x_2$

$\frac{\partial}{\partial \lambda} L = 0$: $\quad x_1 + x_2 - 1 = 0$

$\Rightarrow \boxed{x_1 = x_2 = \frac{1}{2}}$

Now consider inequality constraint(s):

$g(\underline{x}) \geq 0 \qquad \left(\tilde{g}(x) \leq 0 \Leftrightarrow \underbrace{-\tilde{g}(x)}_{g(x)} \geq 0 \right)$

Task: maximize $f(\underline{x})$ s.t. $g(\underline{x}) \geq 0$

Two cases to be distinguished:

(A) constrained stationary point $\underline{x}^*$ with $g(\underline{x}^*) > 0$:

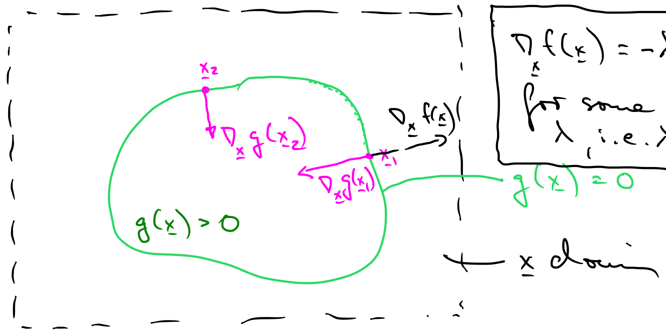
$\Rightarrow$ constraint is inactive and we can choose $\lambda = 0$

(B) the solution $\underline{x}^*$ lies on the boundary of $g(\underline{x}) \geq 0$:

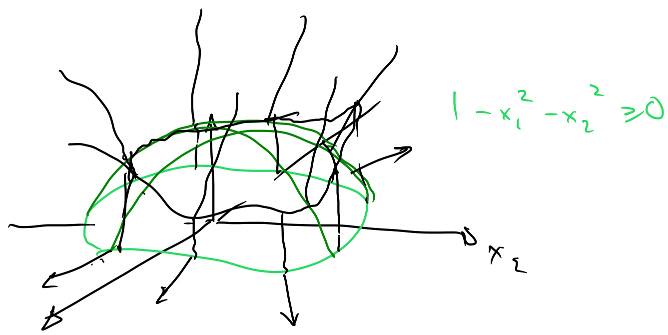
$\Rightarrow$ constraint is active and $\lambda \neq 0$

however: sign of λ is crucial to obtain a maximum or a minimum of $f(\underline{x})$

Illustration: $\max f(\underline{x})$ s.t. $g(\underline{x}) \geq 0$



$\nabla_{\underline{x}} f(\underline{x}) = -\lambda \nabla_{\underline{x}} g(\underline{x})$ for some positive λ, i.e. $\lambda > 0$



In both cases, we have

$$\lambda g(\underline{x}) = 0$$

In total, we obtain for a constrained optimization $\max f(\underline{x})$ s.t. $g(\underline{x}) \geq 0$

with $\quad L(\underline{x},\lambda) = f(\underline{x}) + \lambda g(\underline{x})$

the conditions:

$$\left.\begin{array}{r} g(\underline{x}) \geq 0 \\ \lambda \geq 0 \\ \lambda g(\underline{x}) = 0 \end{array}\right\} \text{Karush-Kuhn-Tucker (KKT) conditions}$$

Special case: if we want to MINIMIZE (rather than MAXIMIZE) $f(\underline{x})$ s.t. $g(\underline{x}) \geq 0$

$\Rightarrow$ we need to have $\nabla_{\underline{x}} f(\underline{x}) = \lambda g(\underline{x})$ with $\lambda \geq 0$ and thus

$$L(\underline{x},\lambda) = f(\underline{x}) - \lambda g(\underline{x})$$

Extension to multiple inequality & equality constraints: $g_j(\underline{x}) = 0$ for $j = 1, \dots, J$ and

$h_k(\underline{x}) \geq 0 \quad$ " $\quad k = 1, \dots, K$

max $f(\underline{x})$ s.t.

$\Rightarrow$ define

$$L(\underline{x}, \lambda_1, \dots, \lambda_J, \mu_1, \dots, \mu_K)$$
$$= f(\underline{x}) + \sum_{j=1}^{J} \lambda_j g_j(\underline{x}) + \sum_{k=1}^{K} \mu_k h_k(\underline{x})$$

s.t. $\mu_k \geq 0$ and $\mu_k h_k(\underline{x}) = 0$ for $k = 1, \dots, K$.

$h_k(\underline{x}) \geq 0 \quad$ " $\quad$ "