# Convex optimization

Virgil Pavlu October 20, 2008

## 1 Convex and differentiable

Say $f$ is convex and differentiable
-then for any $w_1, w_2$ we have

$$f(\frac{w_1 + w_2}{2}) < \frac{1}{2}(f(w_1) + f(w_2))$$

or generalizing

$$f(\lambda w_1 + (1-\lambda)w_2) < \lambda f(w_1) + (1-\lambda)f(w_2)$$

and generalizing on more than two variables (for a set of $\lambda$-s that form a distribution):

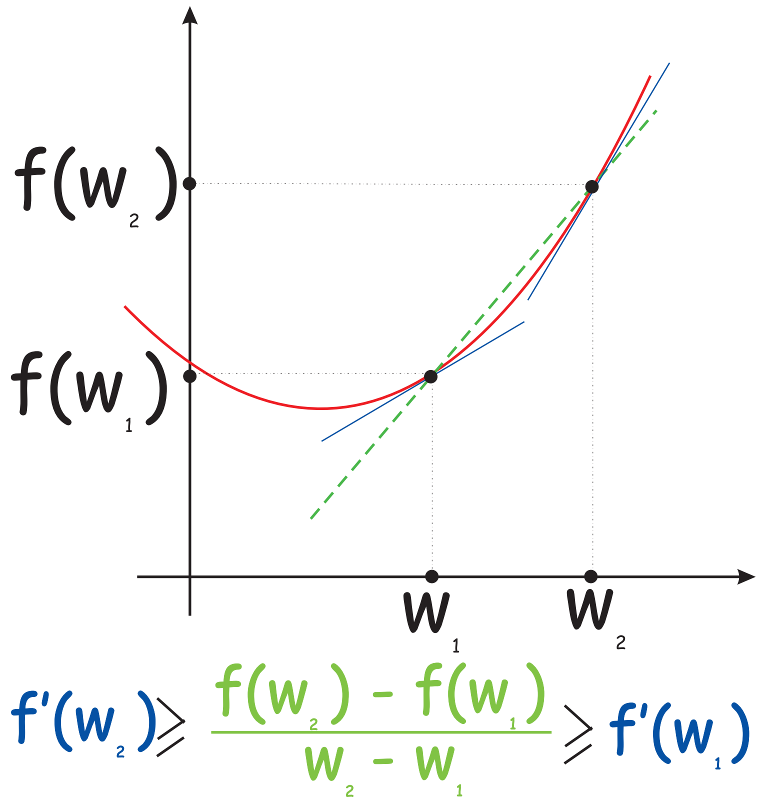
$$f(\sum_i \lambda_i w_i) < \sum_i \lambda_i f(w_i)$$

- Also for any $w_1$,$w_2$ we have (compare the slopes on the figure)

$$f\prime(w_2)(w_2 - w_1) \geq f(w_2) - f(w_1) \geq f\prime(w_1)(w_2 - w_1)$$

There exists $w = \lambda w_1 + (1-\lambda)w_2, \lambda \in [0,1]$ such that

$$f(w_2) - f(w_1) = f\prime(w)(w_2 - w_1)$$




## 2 Constrained Optimization Examples

# constrained optimization

given convex functions $f, g_1, g_2, ...., g_k, h_1, h_2, ...., h_m$ on convex set $X$, the problem

minimize $f(\mathbf{w})$
subject to
$g_i(\mathbf{w}) \leq 0$ ,for all $i$
$h_j(\mathbf{w}) = 0$ ,for all $j$

has as its solution a convex set. If $f$ is strict convex the solution is unique (if exists)

we will assume all the good things one can imagine about functions $f, g_1, g_2, ...., g_k, h_1, h_2, ...., h_m$ like convexity, differentiability etc.That will still not be enough though....

Given that we can write an equality $h_j(\mathbf{w}) = 0$ as two inequality constraints, $h_j(\mathbf{w}) \leq 0$ and $h_j(\mathbf{w}) \geq 0$, we will keep only the inequality constraints.

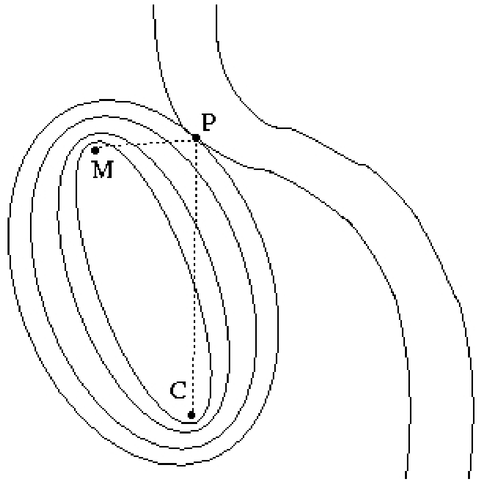



Figure 1: Constrained optimization. Mary has a date with Cal; she wants to get there as soon as possible, but she has to stop by the river first. The optimal route can be seen as follows: all the routes of cost C (fixed) are ellipses centered in M and C; Mary should take consider the smallest ellipse *tangent* to the river. Such an ellipse has the property that the tangent line (differential) in P for both the objective (route) and the constraint (river) have the same direction, there fore the two differentials are a proportional vectors. The proportionality constants are the Lagrangian Multipliers.

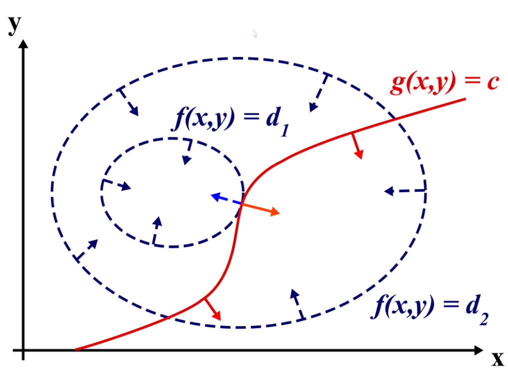



Figure 2: Constrained optimization

## 3 Lagrangian multipliers

We distinguish two types of constraints:
- active : the solution will have $g_i(\mathbf{w}) = 0$
- inactive : the solution will have $g_i(\mathbf{w}) < 0$

Suppose for now (to make things easier) that we know what constraints are going to be active, and so we can write them with equality

lagrange multipliers for equality constraints

let $\Upsilon$ be the feasible region $\Upsilon = \{\mathbf{w} | h_j(\mathbf{w}) = 0 \ \forall j\}$
We assign a lagrangian multiplier to every constraint. So if there are $n$ constraints, we introduce $n$ variables

$\beta = \beta_1, \beta_2, ..., \beta_n)$; The Lagrangian is

$$L(\mathbf{w}, \beta) = f(\mathbf{w}) + \sum_j \beta_j h_j(\mathbf{w})$$

on $\Upsilon$ we have $L(\mathbf{w}, \beta) = f(\mathbf{w})$ and so our problem can be written as
minimize $L(\mathbf{w}, \beta)$
subject to $h_j(\mathbf{w}) = \delta_{\beta_j} L(\mathbf{w}, \beta) = 0$
**Lagrange theorem** nec[essary] and suff[icient] conditions for a point $\widetilde{\mathbf{w}}$ to be an optimum (ie a solution for the problem above) are the existence of $\widetilde{\beta}$ such that

$$\delta_{\mathbf{w}} L(\widetilde{\mathbf{w}}, \widetilde{\beta}) = 0 \textbf{ ; } \delta_{\beta_j} L(\widetilde{\mathbf{w}}, \widetilde{\beta}) = 0$$

**Example.** Find the minimum of the function

$$f(x, y) = (x-1)^2 + (y-2)^2$$

subject to $g(x, y) = 2x + y = 0$

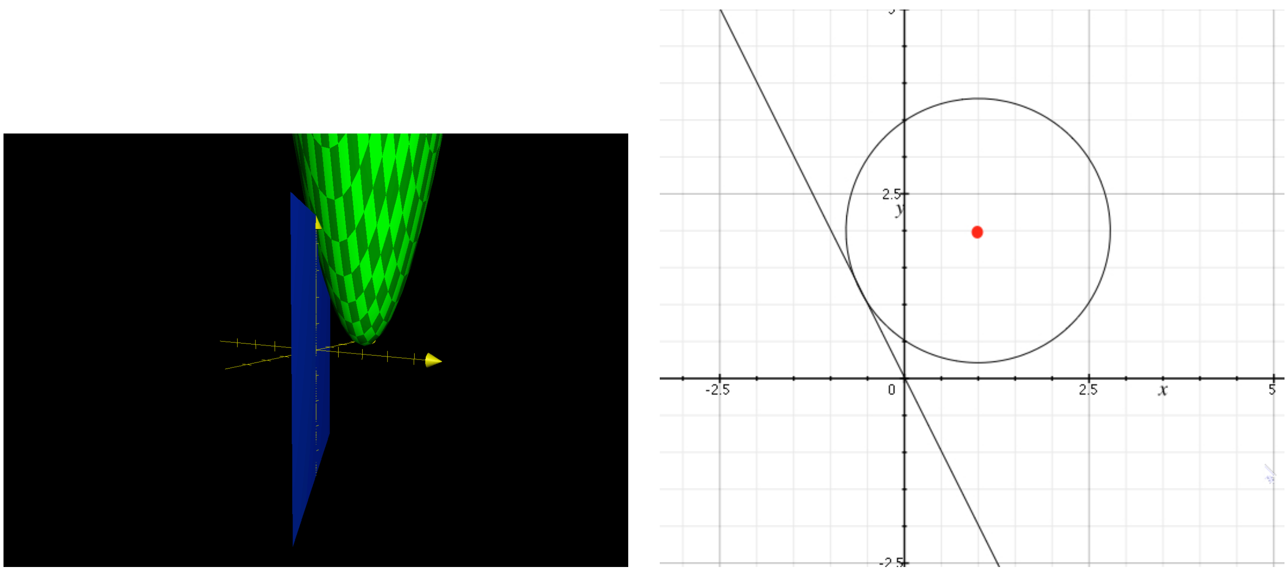



Figure 3: $f$ constrained by $g$

$$L(x, y, \alpha) = (x-1)^2 + (y-2)^2 + \alpha(2x + y)$$

$$\begin{aligned}
\nabla_x L = 2(x-1) + 2\alpha = 0; x = -\alpha + 1 \\
\nabla_y L = 2(y-2) + \alpha = 0; y = \frac{-\alpha + 4}{2} \\
\nabla_\alpha L = 2x + y = 0; -2\alpha + 2 + \frac{-\alpha + 4}{2} = 0; \alpha = 8/5. \\
x = -3/5; y = 6/5
\end{aligned}$$

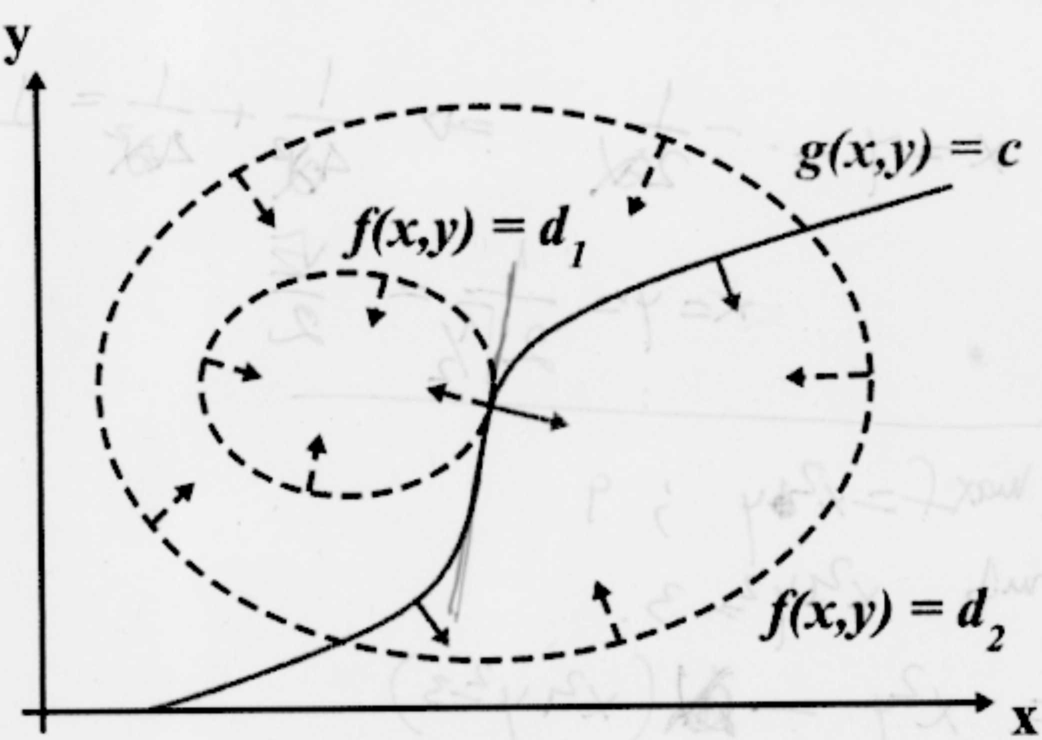
Verification : the center of the circle is at (1,2); the radius vector to the line is (1,2) - (x,y) = (2/5, 4/5) . This vector should be perpendicular on the constraint line.

For inactive constraints, the same exact strategy apply ( they are featured in the lagrangian formula) but their lagrangian multipliers will end up being zero ( KKT theorem) max $f(x,y)$

subject to $g(x,y)=c$

$\lambda$ = Lagrange multipher

$L(x,y,\lambda) = f(x,y) \pm \lambda\,[g(x,y) - c]$




when f contour touces constraint g the two gradients have the same tangent

$\Rightarrow$ gradients are parallel, so there is a proportionality ration $\lambda$ = Lag. mult.

$$\frac{\partial f}{\partial x,y} = +\lambda \frac{\partial g}{\partial x,y}$$

Lagrangian: $f(x,y) + \sum \alpha\, g(x,y)$ one $\alpha$ per constraint

(Lagrange Th): NEC + SUF conditions for $w=(x,y)$ to be OPTIMUM are $\partial_w L(w,\alpha)=0$ ; $\partial_\alpha L(w,\alpha)=0$

Ex: max $f = x+y$
subject to $x^2+y^2=1$

$$L = \underbrace{x+y}_{f} - \lambda(x^2+y^2-1)$$

$$\frac{\partial L}{\partial x} = 1-2\lambda x \quad \Big| \quad \frac{\partial L}{\partial y} = 1-2\lambda y \quad \Big| \quad \frac{\partial L}{\partial \lambda} = x^2+y^2-1$$

$$x = y = -\frac{1}{2\lambda} \Rightarrow \frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} = 1 \Rightarrow \lambda^2 = \frac{1}{2}$$

$$\lambda = \pm\sqrt{1/2}$$

$$x = y = \frac{1}{2\sqrt{1/2}} = \frac{\sqrt{2}}{2}$$

---

Ex: max $f = x^2y$ ; 9
sub $x^2+y^2=3$.

$$L = x^2y - \lambda(x^2+y^2-3)$$

$$\frac{\partial L}{\partial x} = 2xy - 2\lambda x = 0 \Rightarrow x=0 \text{ OR } y=\lambda$$

$x=0 \Downarrow y=\sqrt{3}$, $f=0$

$y=\lambda \Downarrow x^2=2y^2 \Downarrow 3y^2=3 \Rightarrow y=\pm 1$, $x=\pm\sqrt{2}$

$$\frac{\partial L}{\partial y} = x^2 - 2\lambda y = 0$$

$$\frac{\partial L}{\partial \lambda} = x^2+y^2-3 = 0.$$

$f_{max} = 2$

# 4 Kuhn-Tucker Saddle point conditions

## saddle point

minimize $f(\mathbf{w})$
subject to $g_i(\mathbf{w}) \leq 0$ ,for all $i$

and the lagrangian function

$$L(\mathbf{w}, \alpha) = f(\mathbf{w}) + \sum_i \alpha_i g_i(\mathbf{w})$$

$(\widetilde{\mathbf{w}}, \widetilde{\alpha})$ with $\widetilde{\alpha}_i \geq 0$ is **saddle point** if $\forall(\mathbf{w}, \alpha), \alpha_i \geq 0$

$$L(\widetilde{\mathbf{w}}, \alpha) \leq L(\widetilde{\mathbf{w}}, \widetilde{\alpha}) \leq L(\mathbf{w}, \widetilde{\alpha})$$

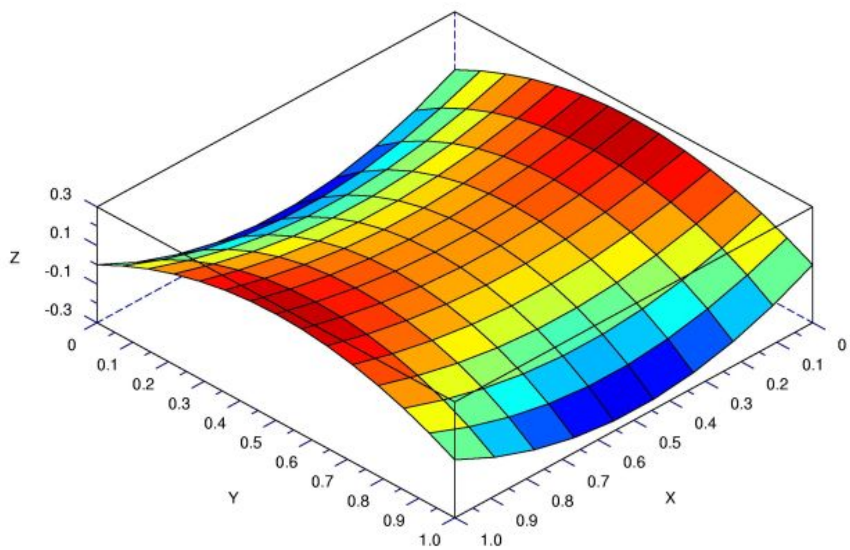
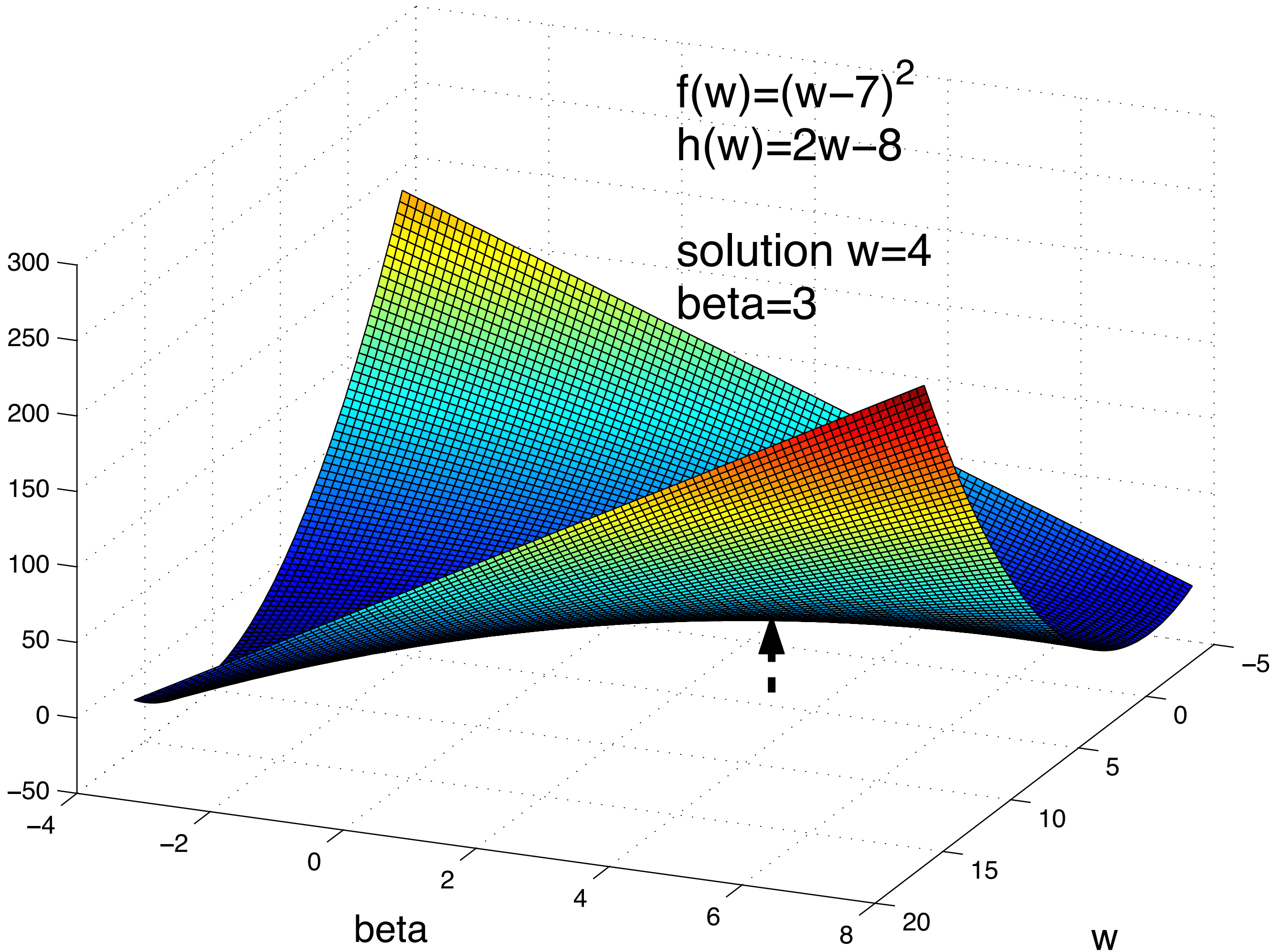



Figure 4: Saddle



f(w)=(w−7)^2
h(w)=2w−8
solution w=4
beta=3
300
250
200
150
100
50
0
−50
−4
−2
0
2
4
6
8
beta
−5
0
5
10
15
20
w

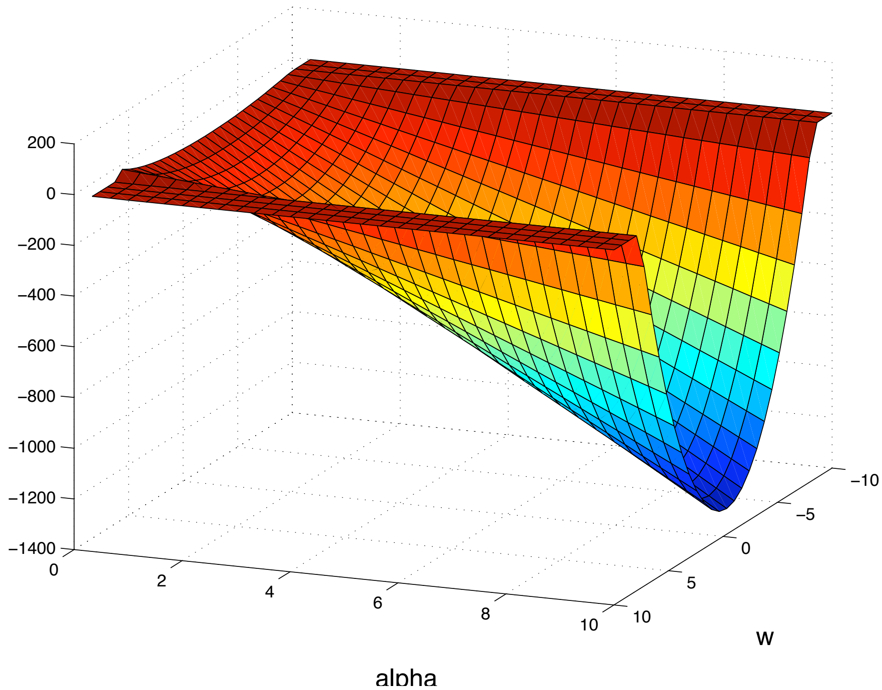



Figure 5: Saddle, linear $\alpha$

## 5 Karush-Kun-Tucker for differentiable, convex problems

# Karush Kuhn Tucker theorem

minimize $f(\mathbf{w})$
subject to $g_i(\mathbf{w}) \leq 0$ ,for all $i$
were $g_i$ are **qualified constraints**

define the lagrangian function

$$L(\mathbf{w}, \alpha) = f(\mathbf{w}) + \sum_i \alpha_i g_i(\mathbf{w})$$

**KKT theorem** nec and suff conditions for a point $\widetilde{\mathbf{w}}$ to be a solution for the optimization problem are the existence of $\widetilde{\alpha}$ such that

$$\begin{aligned}
\nabla_{\mathbf{w}} L(\widetilde{\mathbf{w}}, \widetilde{\alpha}) &= 0 \\
\nabla_{\alpha} L(\widetilde{\mathbf{w}}, \widetilde{\alpha}) &= 0 \\
\widetilde{\alpha}_i g_i(\widetilde{\mathbf{w}}) &= 0 \\
g_i(\widetilde{\mathbf{w}}) &\leq 0 \\
\widetilde{\alpha}_i &\geq 0
\end{aligned}$$

## 6 The dual problem

# duality

$$f(\mathbf{w}) \geq f(\widetilde{\mathbf{w}}) \geq f(\mathbf{w}) + \sum_{i=1}^{k} \alpha_i g_i(\mathbf{w})$$

**dual maximization problem** :
maximize $L(\mathbf{w}, \alpha) = f(\mathbf{w}) + \sum_{i=1}^{k} \alpha_i g_i(\mathbf{w})$
subject to $\alpha \geq 0$ **;** $\delta_{\mathbf{w}} L(\mathbf{w}, \alpha) = 0$

**OR**
set $\theta(\alpha) = \inf_{\mathbf{w}} L(\mathbf{w}, \alpha)$
maximize $\theta(\alpha)$
subject to $\alpha \geq 0$

the primal and dual problem have the same solution if the KKT gap can be vanished

# 7 Interior point methods