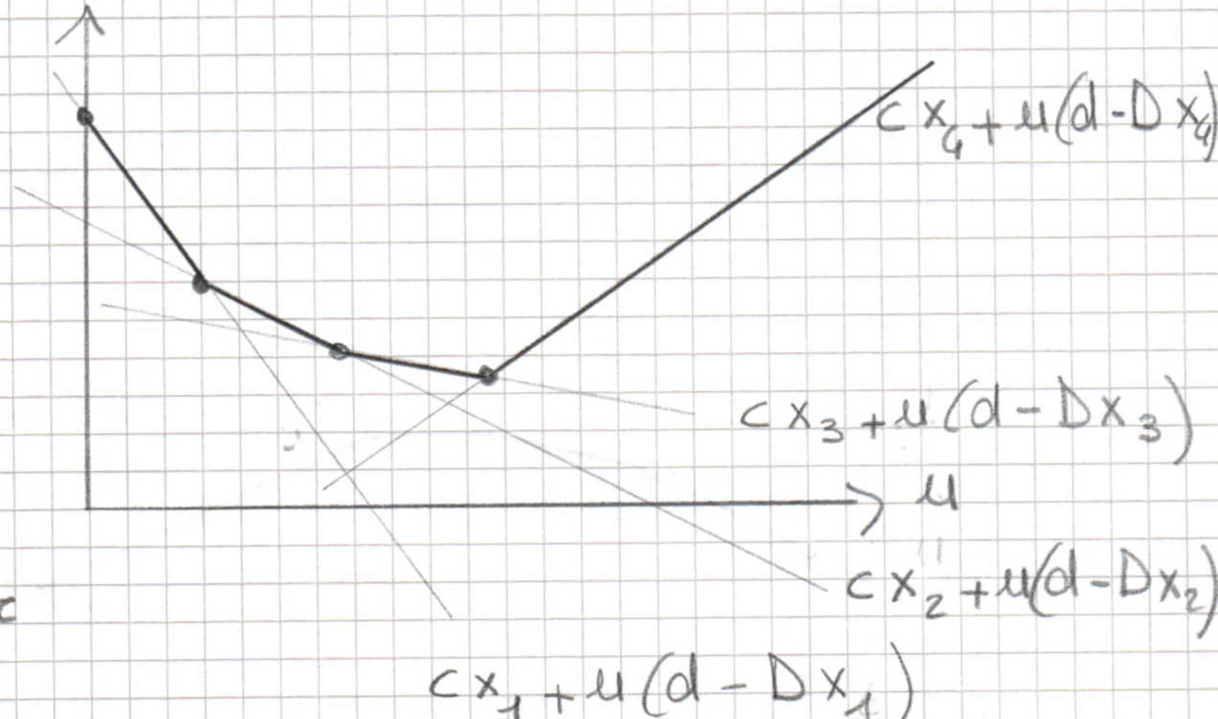
Solving the Lagrangian Dual

(1)

$$w_{LD} = \min_{u \geq 0} z(u)$$

$$= \min_{u \geq 0} \Big\{ \underbrace{\max_{t=1,..T} cx_t + u(d - Dx_t)}_{z(u)} \Big\}$$

i.e. the Lagrangian Dual consists in minimizing a piecewise linear convex, but nondifferentiable, function $z(u)$:



Case of a single multiplier u and $X = \{x_1, x_2, x_3, x_4\}$

How to minimize $z(u)$ over $u \geq 0$?

Solving the corresponding LP is not easy, since it has many (depending on T) constraints : <u>constraint generation approach</u>

<u>an example</u> : (CSP)

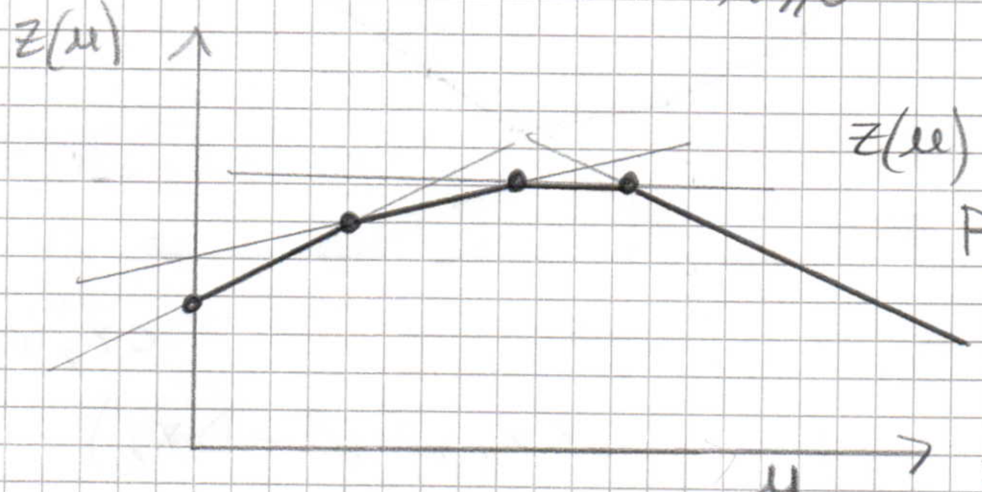
note that if (P) $\quad z = \min cx$

$$Dx \leq d$$
$$x \in X$$

then $(P_u) \quad z(u) = \min cx + u(Dx - d)$

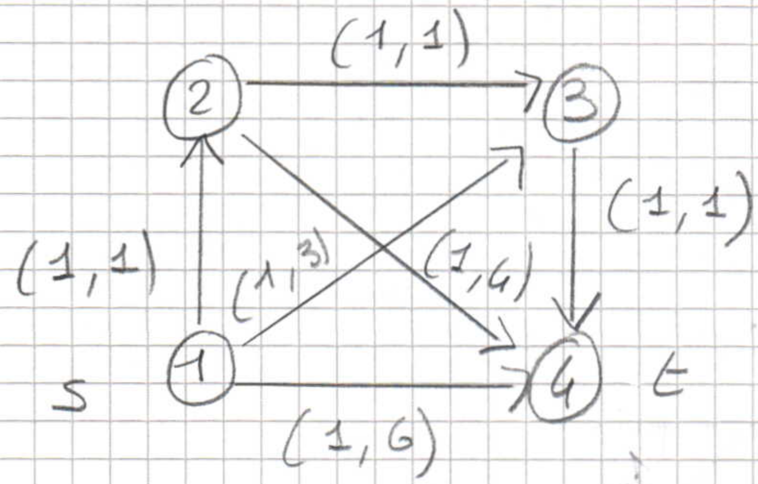
for $u \geq 0$ $\qquad x \in X$

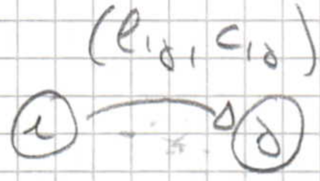
and (LD) $\quad w_{LD} = \max\limits_{u \geq 0} z(u)$



$z(u)$ is now a piecewise linear <u>concave</u> function, to be <u>maximized</u>

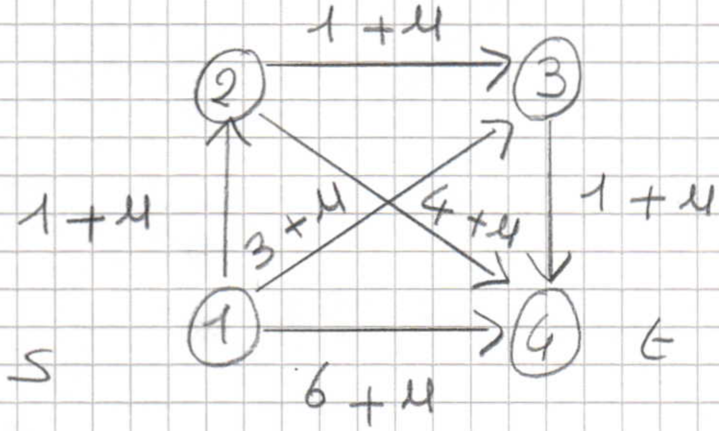
(CSP cont)





$L = 2$

consider the previous Lagrangian relaxation



modified costs

$c_{ij} + l_{ij} u$

- $z(u) = -Lu +$ "shortest path cost w.r.t. $\{c_{ij} + l_{ij} u\}$"

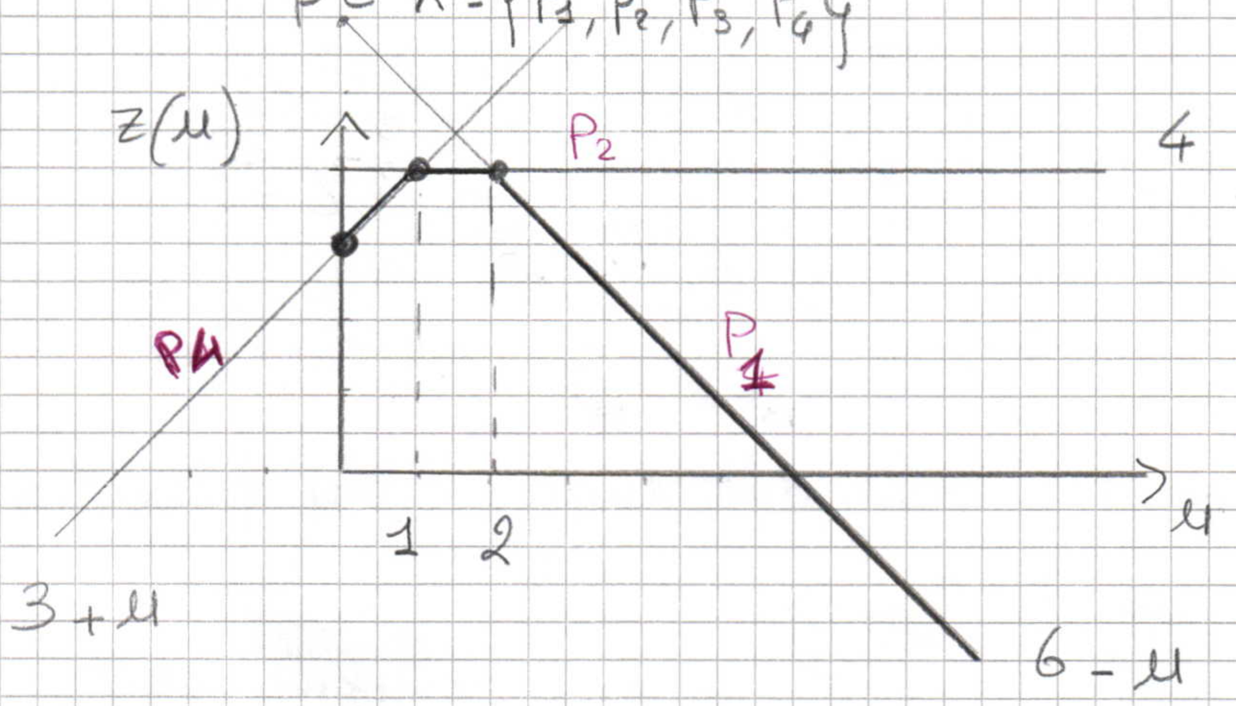
  $-Lu = -2u$

- X is composed of 4 solutions (4 = T paths):

	modified cost	$\boxed{-2u}$
$P_1 = (1, 4)$	$6 + u$	$6 - u$
$P_2 = (1, 3, 4)$	$4 + 2u$	4
$P_3 = (1\ 2\ 4)$	$5 + 2u$	5
$P_4 = (1\ 2\ 3\ 4)$	$3 + 3u$	$3 + u$

$$z(u) = \min \{ 6 - u,\ 4,\ \cancel{5},\ 3 + u \}$$
dominated

$p \in X = \{P_1, P_2, P_3, P_4\}$



* Piecewise linear concave function to be maximize over $u \geq 0$ *

Geometrically : the maximum (best) Lagrangian bound is 4, obtained for $u \in [1\ 2]$;

for $u = 1$; P_4 and P_2 are optimal Lagrangian solutions

for $u \in (0, 1)$; P_2 is optimal Lagr. solution

for $u = 2$; P_2 and P_1 are optimal Lagr. solutions

How to maximize $z(u)$?

(for minimization optimization problems, such as CSP)

Constraint generation approach

let $B \subseteq X$

$$z_B(u) = \min_{x \in B} cx + u(Dx - d)$$

NB: $z_B(u)$ dominates $z(u)$ "from above"

do MASTER PROBLEM

maximize over $u \geq 0$ $z_B(u)$; let z^*_B the maximum value and u^*_B the optimal solution; NB: $z^*_B \geq w_{LD}$

SEPARATION (add a constraint)

compute $z(u^*_B)$, and let x^*_B an optimal solution;

NB: $z(u^*_B)$ is a lower bound to w_{LD} (which is the maximum lagrangian bound); $z(u^*_B) \leq w_{LD}$

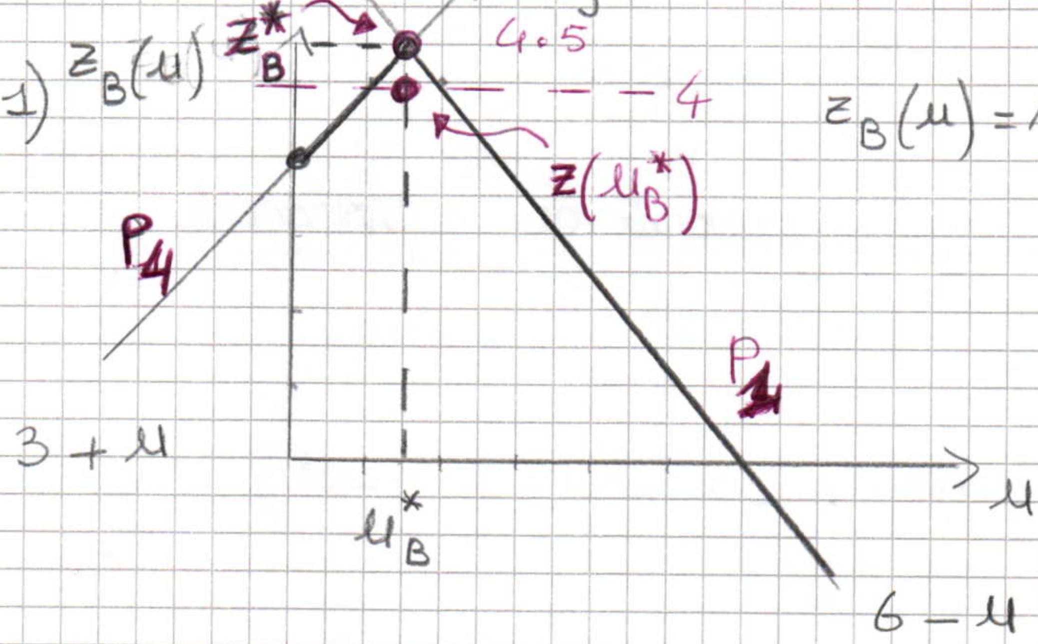
$B := B \cup \{x^*_B\}$ NB: $x^*_B \in X$

while $z(u^*_B) < z^*_B$

example (CSP cont)

Let $B = \{P_1, P_4\}$

1)



$z_B(u) = \min\{6-u, 3+u\}$
$P_4 \qquad P_1$

- maximize $z_B(u)$ over $u \geq 0$ ◄— LP **Master problem**

  $u_B^* = 1.5 \qquad z_B^* = 4.5$

- compute $z(1.5)$: $z(1.5) = 4$,

  found in correspondance to P_2

  < Separation : add $\eta \leq 4$ >

- $B := \{P_1, P_4\} \cup \{P_2\} = \{P_1, P_2, P_4\}$

Since $z(1.5) = 4 < z_B^* = 4.5$

we iterate (with a refined approximation)

2) now $z_B(u) = z(u)$, and after its maximization the algorithm STOP

<u>Number of iterations</u> : $\leq |X|$

<u>Computational efficiency</u>: avoid to consider solutions which are not necessary to compute w_{LD}.

<u>example</u> : start with $B = \{P_2, P_4\}$: in what cases the line corresponding to P_1 is not generated?

<u>example</u> LP corresponding to $B = \{P_1, P_4\}$

$z_B^* = \max \eta$

<u>Master problem</u>

$\eta \leq 6 - \mu$

$\eta \leq 3 + \mu$

$\mu \geq 0$

<u>instead of</u>

$w_{LD} = \max \eta$

$\eta \leq 6 - \mu$

$\eta \leq 3 + \mu$

$\boxed{\eta \leq 4}$

$\mu \geq 0$

The separation adds $\eta \leq 4$

relaxation of the Lagrangian Dual!

(2) An alternative approach (without using LP solvers) is to use a <u>Subgradient Algorithm</u>, which is designed to minimize a piecewise linear convex function (or maximize a piecewise linear concave function).

- Consider the Lagrangian Dual formulation:

$$w_{LD} = \min_{u \geq 0} \{ \underbrace{\max_{t=1,..T} c x_t + u(d - D x_t)}_{z(u)} \}$$

- As already observed, $z(u)$ is piecewise linear convex, but generally it is not differentiable. Therefore, it is possible to use a Subgradient Algorithm.

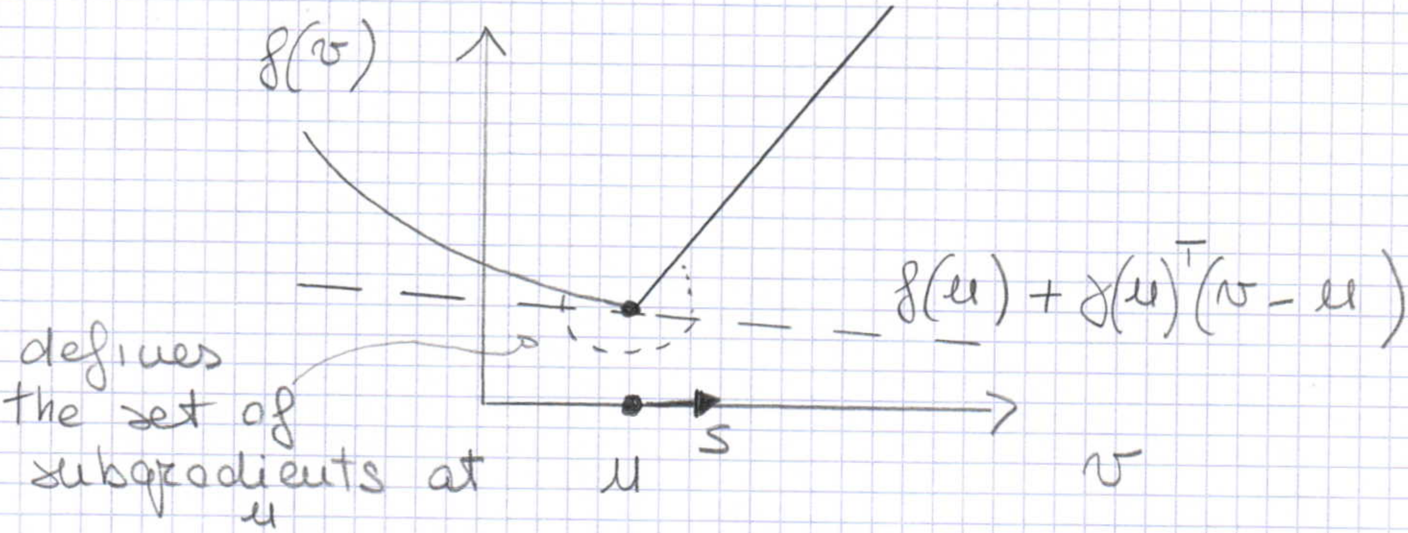
A subgradient is a generalization of gradient:

<u>Definition</u>: Given a convex function $f: \mathbb{R}^m \to \mathbb{R}$, <u>a subgradient of f at u</u> is a vector $\gamma(u) \in \mathbb{R}^m$ such that:

$$f(v) \geq f(u) + \gamma(u)^T (v - u) \quad \forall v \in \mathbb{R}^m$$

If f is a continuously differentiable convex function, then $\gamma(u) = \nabla f(u) = \left(\frac{\partial f}{\partial u_1}, \dots \frac{\partial f}{u_m} \right)$ is the gradient of f at u.

<u>example</u>



<u>Graphically</u>, $\gamma(u)$ is simply the

slope of a <u>supporting hyperplane</u> to the graph of f at u; if f is differentiable at u, the only such plane is the tangent plane, and so $\gamma(u) = \nabla f(u)$.

<u>Property</u> 0 is a subgradient of f at u if and only if u minimizes f.

<u>Proof</u>:

0 is subgradient at u if and only if

$$f(v) \geq f(u) + 0\,(v - u) \quad \forall v, \quad \text{i.e.}$$

$$f(v) \geq f(u) \quad \forall v. \quad \blacksquare$$

<u>Definition</u>: The <u>(one-sided) directional derivative of f at u in the direction s</u> is

$$Df(u; s) = \lim_{\lambda \to 0+} \frac{f(u + \lambda s) - f(u)}{\lambda}$$

So, if $Df(u;s) > 0$, then we increase $f(u)$ by performing a step along s.

In particular:

Theorem: u minimizes f if and only if $Df(u;s) \geq 0$ $\forall$ directions s

What is the relationship between directional derivates and subgradients:

Theorem:

$$Df(u;s) = \max \{ \gamma(u)^T s : \gamma(u) \text{ subgradient of } f \text{ at } u \}$$

No proof

See the figure: if s is the direction in the figure ($s = (1)$), then

$Df(u;s)$ is the right-derivative of f at u, which is the slope of the supporting line at u turned counterclockwise as far as possible: i.e. $\max\{\gamma(u) : \gamma(u)$ subgradient at $u\}$.

Similarly, if $s=(-1)$, $Df(u;s) =$

$= \max\{\gamma(u) s : \gamma(u)$ subgradient at $u\} =$

$= \max\{-\gamma(u) :$ " $\} = -\min\{\gamma(u) :$

$\gamma(u)$ subgradient at $u\}$.

Obs: if f is differentiable at u:

$$Df(u;s) = \underbrace{\nabla f(u)}_{\text{unique subgradient at } u} \cdot s$$

Algorithmic implications: in order to minimize f, choose the direction $s = -\gamma(u)$, since $\gamma(u)^T s = -\gamma(u)^T \gamma(u) \leq 0$, and so it is possible that $Df(u;s) < 0$, allowing to decrease $f(u)$ $\rightsquigarrow$ Subgradient Algorithm

Subgradient Algorithm

Minimization of $z(u)$ over $u \geq 0$

Initialization : $u = u^0$.

Iteration K: $\{u = u^K\}$

- Solve the Lagrangian relaxation (P_{u^K}), and find an optimal solution $x(u^K)$ of (P_{u^K}) [evaluation of $z(u)$ at u^K]
- $u^{K+1} := \max\{u^K - \lambda_K (d - Dx(u^K)), 0\}$
- $K := K + 1$

That is: at each iteration the algorithm moves from the current point u^K along the direction opposite to $(d - Dx(u^K))$, which is a subgradient of function $z(u)$ at u^K:

- $\max\{u^K - \lambda_K (d - Dx(u^K)), 0\}$ since u^{K+1} must be ≥ 0;
- the difficulty is to choose the step lengths $\{\lambda_K\}$.

Property: $(d - Dx(u^k))$ is a subgradient of $z(u)$ at u^k.

Proof

$$z(u^k) = \max_{t=1,..T} cx_t + u^k(d - Dx_t)$$

$$= cx(u^k) + u^k(d - Dx(u^k))$$

In fact, $x(u^k)$ is an optimal solution of $P(u^k)$.

For any $v \geq 0$:

$$z(v) = \max_{t=1..T} cx_t + v(d - Dx_t) \geq$$

$$cx(u^k) + v(d - Dx(u^k)).$$

So $$z(v) \geq cx(u^k) + v(d - Dx(u^k)) +$$

$$+ u^k(d - Dx(u^k)) - u^k(d - Dx(u^k))$$

$$= z(u^k) + \underbrace{(d - Dx(u^k))}_{\text{subgradient at } u^k}(v - u^k)$$

$\square$

How to choose $\{\lambda_k\}$:

- if they are too small, the algorithm can stay "near" the current point and not converge
- if they are too large, the algorithm may skip the optimal solution

Compromise:

1) $\{\lambda_k\} \to 0$ when $k \to +\infty$
2) $\sum_{j=1}^{k} \lambda_k \to +\infty$ when $k \to +\infty$

These conditions ensure that the Subgradient Algorithm converges to w_{LD} (no proof).

example $\lambda_k = \frac{1}{k}$ $\forall k \geq 1$

example (CSP) cont.

Let $u^0 = 3$

Iteration ($u = 3$)

- Solve (P_3), finding $z(3) = (6-3) = 3$ and the optimal solution $x(3)$ which is the path P_1 (see the figure in 50)

- the complicating constraint $Dx \leq d$ is

$$\sum_{(i,j)\in A} \ell_{ij} x_{ij} \leq \underset{\displaystyle 2}{\overset{}{L}} \quad ;$$

(CSP) is a minimization optimization problem, and we want to maximize $z(u)$, so we move along $(Dx(u^k) - d)$ which is a supergradient of the concave function $z(u)$ (or simply a subgradient):

$$\sum_{(i,j)\in A} \ell_{ij} x_{ij}(3) - L = \sum_{(i,j)\in P_1 = (1,4)} \ell_{ij} - 2 = 1 - 2 = -1$$

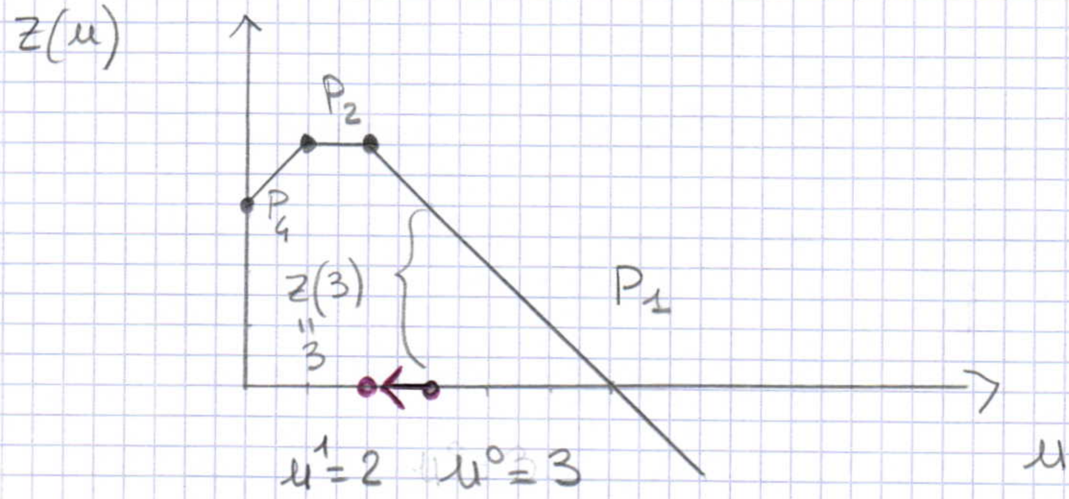
gradient at $u = 3$

- so $u^1 := \max\{u^0 + \lambda_1(-1), 0\} =$

$$= \max\{3 - \lambda_1, 0\}$$

If $\lambda_1 = 1$, then:

$u^1 := 2$



$K := 1$

<u>Iteration</u> ($u = 2$)

- Solve (P_2), finding $z(2) = 4$ and an optimal solution $x(2)$; if $x(2)$ is the path P_2, then $\sum_{(i,j) \in P_2} e_{ij} - 2 = 0$ ↑ <u>Maximum!</u>

- $u^2 = u^1$ and the algorithm STOP

<u>In general</u>: the algorithm is often terminated before w_{LD} is attained.

Exercise:

If (P) $\min cx$

$$Dx \leq d$$
$$x \in X$$

prove that $(Dx(u^k) - d)$ is a supergradient of $z(u)$ at u^k, i.e.

$$z(v) \leq z(u^k) + (Dx(u^k) - d)(v - u^k) \quad \forall v \geq 0$$

recall that $x(u^k)$ denotes an optimal solution of the lagrangian relaxation P_{u^k}.