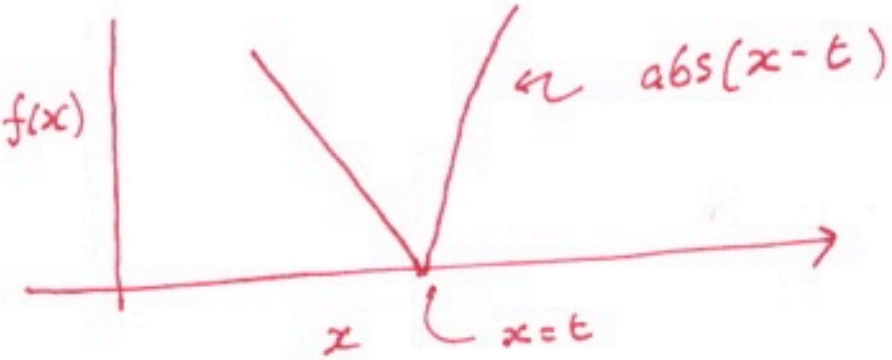
Subgradients

- what if f is continuous, but not everywhere differentiable?



for $x < t$, $\frac{\partial f}{dx} = -1$

$x > t$, $\frac{df}{dx} = 1$

at $x = t$?

one idea

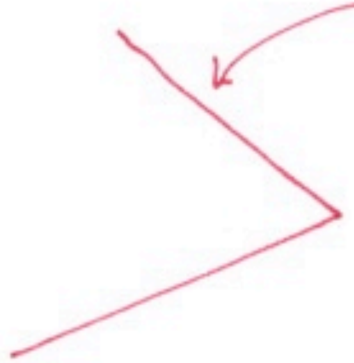
$\partial_x f$ = subgradient of f

$= [-1; 1]$

interval

$$\partial_x f = \begin{cases} \text{gradient when } f \text{ diff} \\ ? \text{ otherwise} \end{cases}$$

This is fairly natural

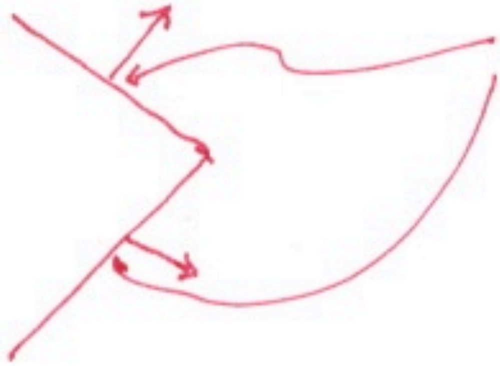


curve $(x(s), y(s))$

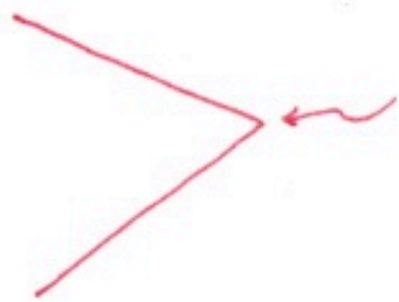
~~normal~~ tangent $\left(\frac{dx}{ds}, \frac{dy}{ds}\right) \equiv (\dot{x}, \dot{y})$

normal

$(\dot{y}, -\dot{x})$



over here, there is a <u>unique</u> normal

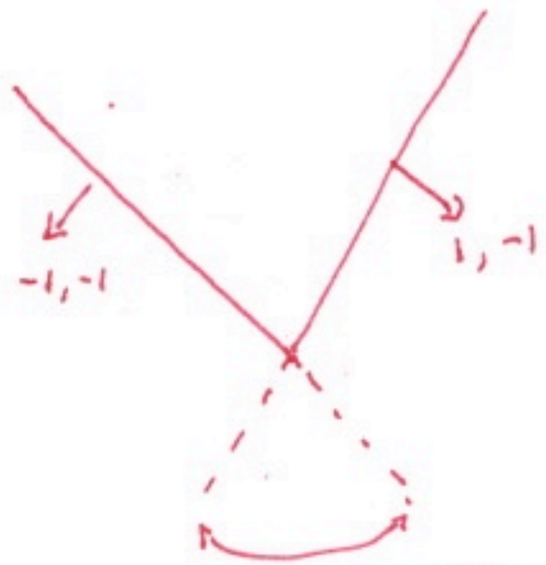


here, there is a <u>family</u> of tangents, and so a <u>family</u> of normals (it's a <u>cone</u>)

$(s, \mathrm{abs}(s-t))$ (our original curve

<u>tangent</u>: $(1, \partial_s \mathrm{abs}(s-t))$

<u>normal</u>: $(\partial_s \mathrm{abs}(s-t), -1) \times$ sign (outward pointing



$(s, -1)$ for $s \in [-1, 1]$

all this works in higher dimensions, (check) just harder to draw.

Two key points

- We now have an improved method to check whether we are at minimum

min, but how do we know?

$0 \in \partial_x f$ ← v. useful in practice:

- all grad descent works with subgradient, too!
  - If you are at x where f is cont but not diff, use any value of subgradient
  - Good for theory, not much significance in practice