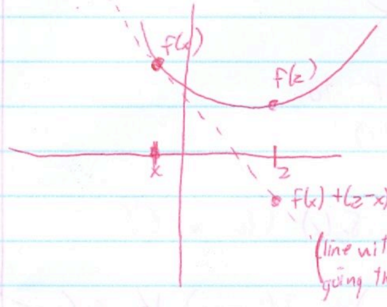
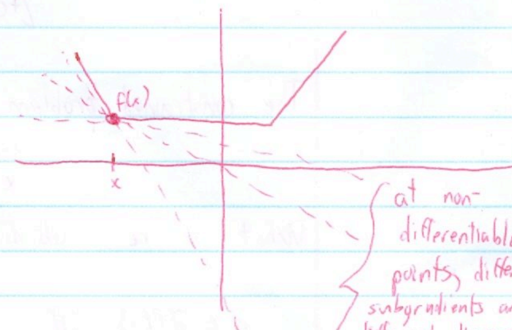
Outline

1. Optimality for Non-smooth Convex Optimization
2. Polar Cones
3. Examples

1. Optimality for Non-Smooth Convex Optimization

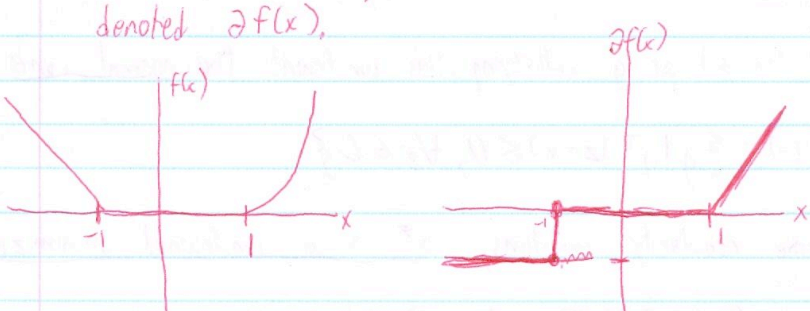
- Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex.
- A vector d is a subgradient of f at x if
$$f(z) \geq f(x) + (z-x)^T d, \quad \forall z \in \mathbb{R}^n$$





at non-differentiable points, different subgradients are different linear underestimators

- The set of all sub-gradients of f at x is the subdifferential, denoted $\partial f(x)$.



- x^* is an unconstrained minimizer iff $0 \in \partial f(x)$

eg. if $0 \in \partial f(x)$, then $f(z) \geq f(x) + (z-x)^T 0, \; \forall z \in \mathbb{R}^n$
$\Rightarrow f(z) \geq f(x), \; \forall z \in \mathbb{R}^n$

Constrained Case

Let $C \subseteq \mathbb{R}^n$ be a non-empty convex set.

~~What about~~

Consider: $\min f(x)$
s.t. $x \in C$

We can handle this using the same optimality condition using an extended real-valued indicator function:

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

The constrained problem is equivalent to:

$$\min_x f(x) + \delta_C(x)$$

What is the sub-differential of $\delta_C(x)$ (for $x \in C$)?

$$g \in \partial f(x) \text{ iff } \delta_C(z) \geq \delta_C(x) + g^T(z-x), \forall z \in C$$

we say $\partial f(x) = \emptyset$ if $x \notin C$

if $z \notin C$, then reduces to $+\infty \geq \delta_C(x) + g^T(z-x)$

but $\delta_C(z) = \delta_C(x) = 0$, so $g \in \partial f(x)$ iff
$g^T(z-x) \leq 0, \forall z \in C$

We call the set of g satisfying this constraint the <u>normal cone</u>

$$N_C(x) = \{ g \mid g^T(z-x) \leq 0, \forall z \in C \}$$

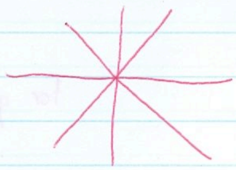
Under some regularity conditions, <u>x^* is a constrained minimizer</u> if

$$0 \in \partial f(x^*) + N_C(x^*)$$

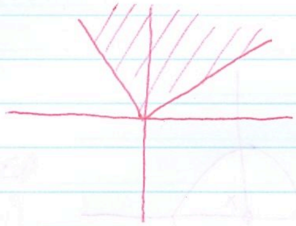
(regularity condition: dom(f) overlaps with C,)
so function isn't $+\infty$ everywhere

2. Polar Cones

- A *cone* is a set C such that $\forall_x \in C$, $\alpha x \in C$ for $\alpha \geq 0$



- A *convex cone* is a set C such that $\forall_{x,y} \in C$, $\alpha x + \beta y \in C$ for $\alpha \geq 0, \beta \geq 0$

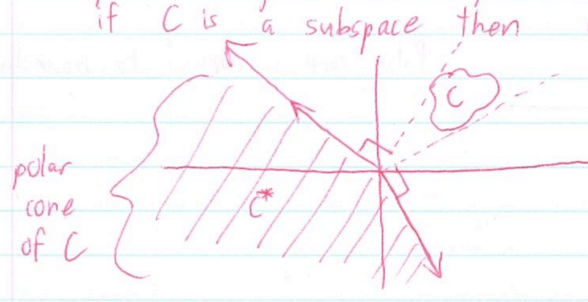


"Closed under non-negative linear combinations"

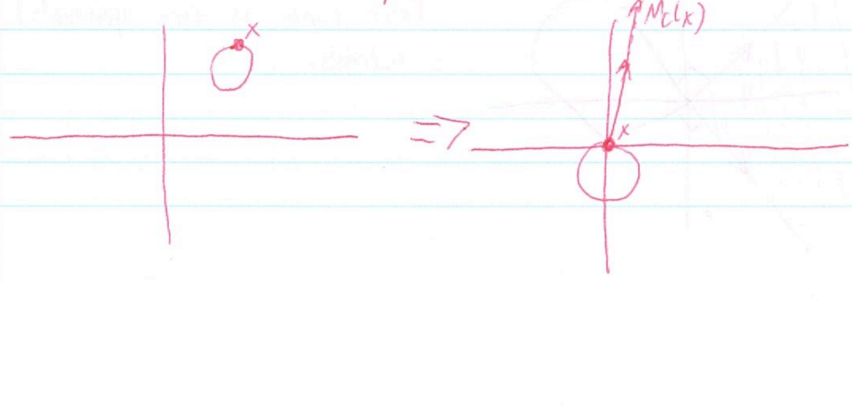
- The *polar cone* of a set C is the convex cone defined as:

$$C^* \triangleq \{y \mid y^T x \leq 0, \ \forall_x \in C\}$$

- The polar cone *generalizes orthogonal complement* of subspaces, if C is a subspace then $C^* = C^\perp$

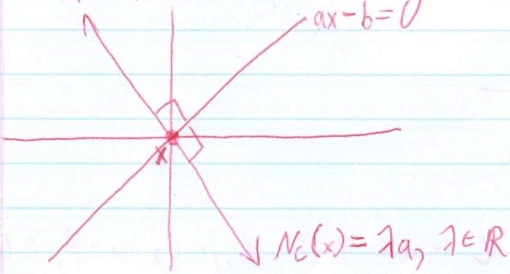


- The *normal cone at x is the polar cone of $C - \{x\}$*



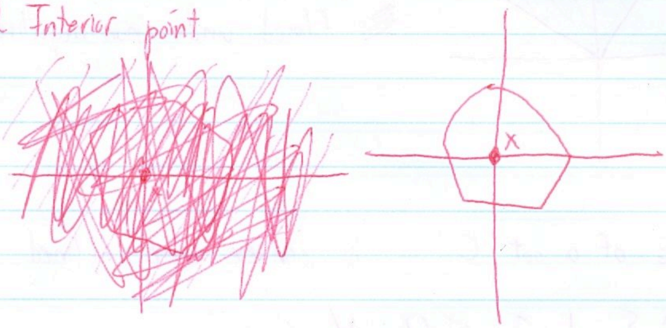
Examples of Polar Cones and Normal Cones

1. Linear Equality constraints



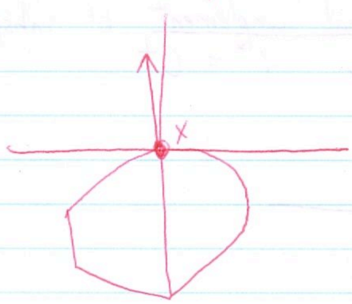
For general linear equality constraint $Ax = b$,
$N_C(x) = R(A^T)$
← row-space of A
(for y to be feasible, it must be in null-space of A)

2 Interior point



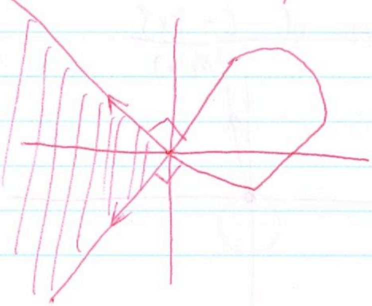
Polar cone is $N_C(x) = \{0\}$

3. Smooth Boundary Point



Polar cone is normal to boundary

4. Non-smooth Boundary Point



Polar cone is cone generated by normals.

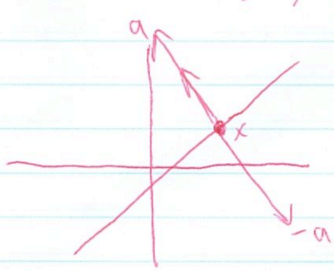
Examples

1. $\min f(x)$ (differentiable, convex)
   s.t. $x \in \mathbb{R}^n$

- normal cone is $\{0\}$, $\partial f(x) = \nabla f(x)$, so we get the usual $\nabla f(x) = 0$

2. $\min f(x)$ (differentiable, convex)
   s.t. $a^T x = b$

- normal cone is $\{\lambda a, \lambda \in \mathbb{R}^n\}$, so we get $\underbrace{\nabla f(x) + \overset{\text{Lagrange multiplier}}{\lambda} a}_{\text{Lagrangian}} = 0$ for some $\lambda \in \mathbb{R}^n$



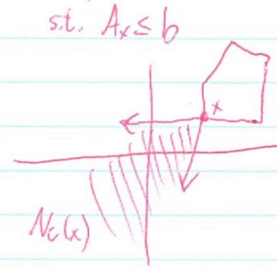
you can make progress if $\nabla f(x)$ is not parallel to a

3. $\min \|x\|_1$
   s.t. $Ax = b$

- normal cone is $\{s \mid s = \sum_i \lambda_i a_i\} = R(A^T)$, $\partial f(x) = \operatorname{sgn}(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \end{cases}$,

so we get ~~$\partial f(x) \in R(A^T)$~~
$0 \in \operatorname{sgn}(x) + R(A^T)$

4. $\min f(x)$ (convex)
   s.t. $Ax \leq b$



(you can make progress if $-\nabla f(x)$ is not in normal cone)

normal cone is $\{s \mid s = \sum_{i \in A(x)} \lambda_i a_i, \lambda_i \geq 0\}$, ($A(x)$: active set)

so we get $\partial f(x) + \sum_{i \in A(x)} \lambda_i a_i$,

for some $\lambda_i \geq 0$.