Lecture 6
Subgradients and subdifferentials

– subgradients and quasigradients
– subdifferentials
– subgradient calculus

Motivation

extend notion of gradient to

– nondifferentiable convex functions
– quasiconvex functions

idea: given x_k, we need to 'rule out' a halfspace at x_k, *i.e.*, find $g \neq 0$ s.t.

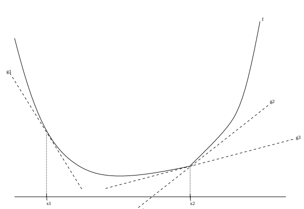
$$g^T(x^\star - x_k) \leq 0$$

– for differentiable fcts, g can be gradient
– but *any* such g will work ...

Subgradient of a convex function

g is a *subgradient* of f at x if

$$f(y) \geq f(x) + g^T(y - x) \quad \text{for all } y$$



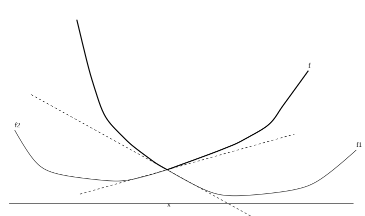
g_2, g_3 are subgradients at x_2; g_1 is a subgradient at x_1

– subgradient gives affine global lower bound on f
– $g^T(y - x) \geq 0 \Longrightarrow f(y) \geq f(x)$
– a convex function f is subdifferentiable (*i.e.*, at least one subgradient exists) at all points in **relintdom** f

Examples

$$f = \max\{f_1, f_2\}$$

with f_1, f_2 convex and differentiable



– $f_1(x_0) > f_2(x_0)$: unique subgradient $g = \nabla f_1(x_0)$
– $f_2(x_0) > f_1(x_0)$: unique subgradient $g = \nabla f_2(x_0)$
– $f_1(x_0) = f_2(x_0)$: subgradients form a line segment

$$[\nabla f_1(x_0), \nabla f_2(x_0)]$$

Subgradient of largest eigenvalue

For a symmetric matrix X, define

$$f(X) = \lambda_{\max}(X)$$

This function is convex (epigraph is)

Let u be a unit-norm eigenvector corresponding to largest eigenvalue, then a subgradient at X_0 is

$$G = uu^T$$

Proof: for every X,

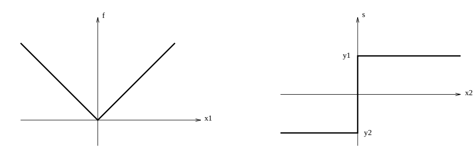
$$\begin{aligned}\lambda_{\max}(X) \geq u^T X u &= u^T X_0 u + (u^T X u - u^T X_0 u) \\ &= \lambda(X_0) + \mathbf{Tr}\, G(X - X_0)\end{aligned}$$

Subdifferentials

set of all subgradients of f at x is called the *subdifferential* of f at x, written $\partial f(x)$

- $\partial f(x)$ is a closed convex set
- $\partial f(x)$ nonempty (if f convex, and finite near x)
- $\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x
- if $\partial f(x) = \{g\}$, then f is differentiable at x and $g = \nabla f(x)$
- in most applications (*e.g.*, ellipsoid method), don't need complete $\partial f(x)$; it is sufficient to find one $g \in \partial f(x)$

example: $f(x) = |x|$



Calculus of subgradients

assumption: all functions are finite near x

- $\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x
- *scaling:* $\partial(\alpha f) = \alpha \partial f$ (if $\alpha > 0$)
- *addition:* $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- *affine transformation of variables:* if $g(x) = f(Ax+b)$, then $\partial g(x) = A^T \partial f(Ax+b)$
- *pointwise maximum:* if $f = \max_{i=1,\ldots,m} f_i$, then

  $$\partial f(x) = \mathbf{Co} \cup \{\partial f_i(x) \mid f_i(x) = f(x)\},$$

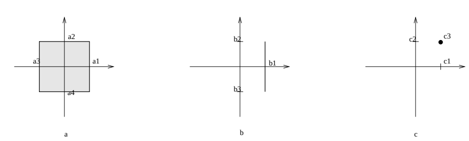
  i.e., convex hull of union of subdifferentials of 'active' functions at x

  special case: if f_i differentiable

  $$\partial f(x) = \mathbf{Co}\, \{\nabla f_i(x) \mid f_i(x) = f(x)\}$$

example

$$f(x) = \|x\|_1 = \max\{s^T x \mid s_i \in \{-1, 1\}\}$$



- *pointwise supremum:* if $f = \sup_{\alpha \in \mathcal{A}} f_\alpha$, then

  $$\partial f_\beta(x) \subseteq \partial f(x)$$

  if $f_\beta(x) = f(x)$ and $\beta \in \alpha$

  (many technical conditions required for equality)

  example

  $$f(x) = \lambda_{\max}(A(x)) = \sup_{\|y\|=1} y^T A(x) y$$

  where $A(x) = A(x)^T = A_0 + x_1 A_1 + \cdots + x_n A_n$

  – $g_y(x) \triangleq y^T A(x) y$ is affine in x, with

  $$\nabla g_y(x) = (y^T A_1 y, \ldots, y^T A_n y)$$

  – hence,

  $$\partial f(x) = \mathbf{Co}\ \{\nabla g_y \mid A(x)y = \lambda_{\max}(A(x))y,\ \|y\| = 1\}$$

  (not hard to verify)

– *minimization:* define $g(y)$ as the optimal value of

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le y_i, \quad i = 1, \dots, m \end{array}$$

(f_i convex; variable x)
from duality (*c.f.*, page 7-17):

$$g(y) \ge g(0) - \sum_{i=1}^{m} \lambda_i^\star y_i$$

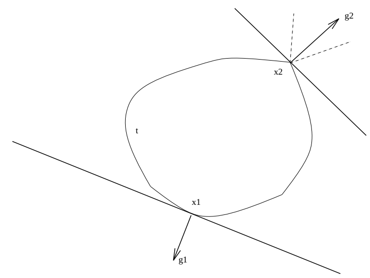
i.e., $-\lambda^\star$ is a subgradient of g at $y = 0$

Subgradients and sublevel sets

g is a subgradient at x if

$$f(y) \ge f(x) + g^T(y - x)$$

hence $f(y) \le f(x) \Longrightarrow g^T(y - x) \le 0$

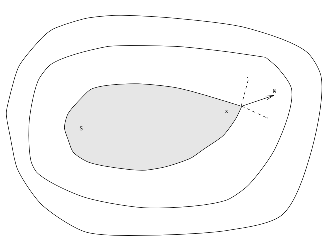


– f differentiable at x_0: $\nabla f(x_0)$ is normal to the sublevel set $\{x \mid f(x) \le f(x_0)\}$
– f nondifferentiable at x_0: subgradient defines a supporting hyperplane to sublevel set through x_0

Quasigradients

if f is quasiconvex, then g is a *quasigradient* if

$$g^T(y - x) \ge 0 \Rightarrow f(y) \ge f(x)$$



– allows us to rule out a halfspace when minimizing f
– quasigradients at x_0 form a cone

Examples

$$f(x) = \frac{a^T x + b}{c^T x + d}, \quad (\textbf{dom}\, f = \{x | c^T x + d > 0\})$$

$g = a - f(x_0)c$ is a quasigradient at x_0

proof: for $c^T x + d > 0$:

$$a^T(x - x_0) \ge f(x_0) c^T (x - x_0) \Longrightarrow f(x) \ge f(x_0)$$

example: degree of $a_1 + a_2 t + \cdots + a_n t^{n-1}$

$$f(a) = \min\{i \mid a_{i+2} = \cdots = a_n = 0\}$$

$g = \text{sign}(a_{k+1}) e_{k+1}$ (with $k = f(a)$) is a quasigradient at $a \ne 0$

proof:

$$g^T(b - a) = \text{sign}(a_{k+1}) b_{k+1} - |a_{k+1}| \ge 0$$

implies $b_{k+1} \ne 0$