Subgradients

Operations Research

Anthony Papavasiliou

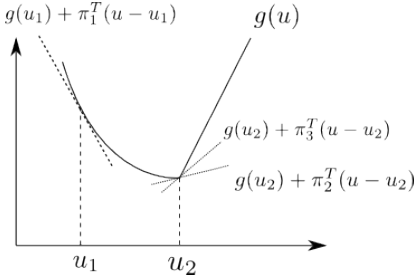
Contents

Table of Contents

Subgradient of a function

π is a subgradient of g (not necessarily convex) at u if

$$g(w) \geq g(u) + \pi^T(w - u) \text{ for all } w$$



π_1 is a subgradient at u_1; π_2, π_3 are subgradients at u_2

The subgradient is a generalization of ...?

- π is a subgradient iff $g(u) + \pi^T(w - u)$ is a global (affine) underestimator of g
- If g is convex and differentiable, $\nabla g(u)$ is a subgradient of g at u

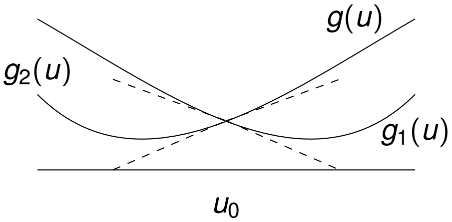
Subgradients come up in two types of algorithms that we will study

- Dual decomposition
- L-shaped method and extensions

(If $g(w) \leq g(u) + \pi^T(w - u)$ for all w, then π is a supergradient)

Example

$g = \max\{g_1, g_2\}$ with g_1, g_2 convex and differentiable



- $g_1(u_0) > g_2(u_0)$: unique subgradient $\pi = \nabla g_1(u_0)$
- $g_2(u_0) > g_1(u_0)$: unique subgradient $\pi = \nabla g_2(u_0)$
- $g_1(u_0) = g_2(u_0)$: subgradients form a line segment $[\nabla g_1(u_0, \nabla g_2(u_0))]$

Subdifferential

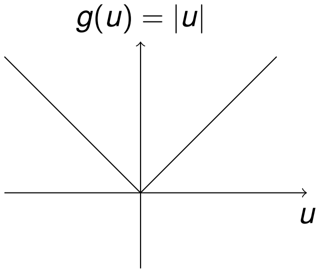
- Set of all subgradients of g at u is called the **subdifferential** of g at u, denoted $\partial g(u)$
- $\partial g(u)$ is a closed convex set

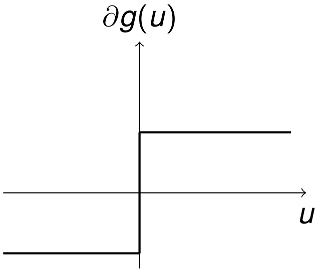
If g is convex

- $\partial g(u)$ is nonempty, for $u \in \text{relint dom } g$
- $\partial g(u) = \{\nabla g(u)\}$, if g is differentiable at u
- If $\partial g(u) = \{\pi\}$, then g is differentiable at u and $\pi = \nabla g(u)$

Example

$g(u) = |u|$





Right hand plot shows $\cup\{(u, \nabla g(u))|u \in \mathbb{R}\}$

Table of Contents

Some Basic Rules

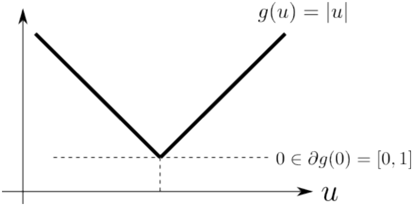
Suppose g is convex

- $\partial g(u) = \{\nabla g(u)\}$ if g is differentiable at u
- Scaling: $\partial(ag) = a\partial g$
- Addition: $\partial(g_1 + g_2) = \partial g_1 + \partial g_2$ (RHS is addition of sets)
- Affine transformation of variables: if $f(u) = g(Au + b)$, then $\partial f(u) = A^T \partial g(Au + b)$
- Finite point wise maximum: if $g = \max_{i=1,\dots,m} g_i$, then

$$\partial g(u) = \text{Co} \cup \{\partial g_i(u) | g_i(u) = g(u)\}$$

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

Example



Consider $g(u) = |u|$, note that

$$\partial g(0) = \text{Co}(\{-1\} \cup \{1\}) = [-1, 1]$$

Optimality Conditions - Unconstrained

Recall for g convex, differentiable,

$$g(u^\star) = \inf_u g(u) \Leftrightarrow 0 = \nabla g(u^\star)$$

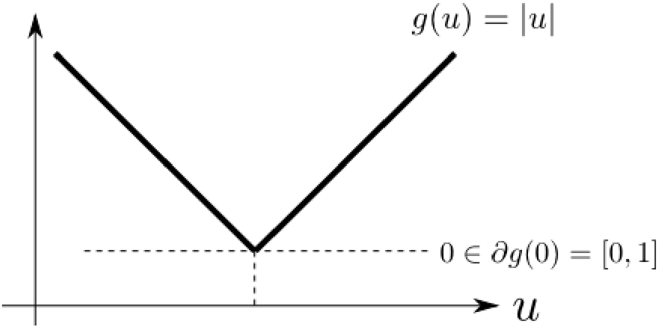
Generalization to non-differentiable convex g

$$g(u^\star) = \inf_u g(u) \Leftrightarrow 0 \in \partial g(u^\star)$$

Proof. By definition

$$g(w) \geq g(u^\star) + 0^T(w - u^\star) \text{ for all } w \Leftrightarrow 0 \in \partial g(u^\star)$$

Example



Parametrizing the Right-Hand Side

Define $c(u)$ as optimal value of

$$c(u) = \min f_0(x)$$
$$f_i(x) \leq u_i, i = 1, \ldots, m$$

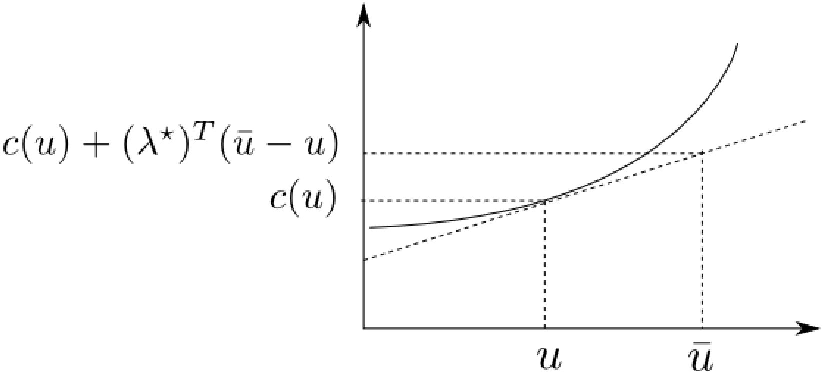
where $x \in \text{dom } f_0$ and f_0, f_i are convex functions

- $c(u)$ is convex
- Suppose strong duality holds and denote $\lambda^\star$ as the maximizer of the dual function

$$\inf_{x \in \text{dom } f_0} (f_0(x) - \lambda^T(f(x) - u))$$

for $\lambda \leq 0$. Then $\lambda^\star \in \partial c(u)$.

Graphical Illustration



Proof: $c(u)$ Is Convex

- Consider any u_1, u_2, denote x_1, x_2 as the respective optimal solutions.
- Consider any $a \in [0, 1]$ and denote x_a as the optimal solution when $au_1 + (1 - a)u_2$ is used as input

- Convexity of $f \Rightarrow f(ax_1 + (1 - a)x_2) \leq au_1 + (1 - a)u_2$ (since $f(x_1) \leq u_1$ and $f(x_2) \leq u_2$)
- Convexity of dom $f_0 \Rightarrow ax_1 + (1 - a)x_2$ is admissible when $au_1 + (1 - a)u_2$ is used as input
- Optimality of x_a with respect to $au_1 + (1 - a)u_2 \Rightarrow$ $f_0(x_a) \leq f_0(ax_1 + (1 - a)x_2)$
- Convexity of $f_0 \Rightarrow$ $c(au_1 + (1 - a)u_2) \leq ac(u_1) + (1 - a)c(u_2)$

Proof: $\lambda^\star$ Is a Subgradient

- Denote $\bar{x}$ as the optimal solution for $\bar{u}$
- Denote $x^\star \in \arg\min_{x \in \mathrm{dom}} (f_0(x) - (\lambda^\star)^T (f(x) - u))$

$c(u) = f_0(x^\star) - (\lambda^\star)^T (f(x^\star) - u) \leq$	strong duality
$f_0(\bar{x}) - (\lambda^\star)^T (f(\bar{x}) - u) =$	definition of $x^\star$
$f_0(\bar{x}) - (\lambda^\star)^T (f(\bar{x}) - \bar{u}) - (\lambda^\star)^T (\bar{u} - u) \leq$	
$f_0(\bar{x}) - (\lambda^\star)^T (\bar{u} - u) =$	since $f(\bar{x}) \leq \bar{u}$, $\lambda^\star \leq 0$
$c(\bar{u}) - (\lambda^\star)^T (\bar{u} - u)$	definition of $\bar{x}$

Example: The Diet Problem

Problem: Choose 3 dishes (x_1, x_2, x_3) so as to satisfy nutrient requirements b_1 and b_2, while minimizing cost (dishes cost 1 \$, 2 \$, and 3 \$ respectively)

Table: The unit of nutrients in each dish.

	Dish 1	Dish 2	Dish 3
Nutrient 1	0.5	4	1
Nutrient 2	2	1	2

$$
\begin{aligned}
z(b) = \quad \min \quad & x_1 + 2x_2 + 3x_3 \\
\text{s.t.} \quad & 0.5x_1 + 4x_2 + x_3 = b_1 \\
& 2x_1 + x_2 + 2x_3 = b_2 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

If $b \geq 0$, then (we showed this in the previous lecture)

$$z(b) = \begin{cases} +\infty, & b_2 > 4b_1 \\ 0.4b_1 + 0.4b_2, & 0.25b_1 \leq b_2 \leq 4b_1 \\ +\infty, & b_2 < 0.25b_1 \end{cases}$$

This is a convex function

Sensitivity

Corollary of previous proposition: if $c(u)$ is differentiable at u, then $\lambda^\star = \nabla c(u)$

$\Rightarrow \lambda_i$ is equal to the *sensitivity* of $c(u)$ to a marginal change in the right-hand-side of the constraint corresponding to λ_i

Example: The Diet Problem - Sensitivity

Consider the diet problem with $b_1 = 1$ and $b_2 = 2$

Show that $\pi_1^\star = 0.4$ and $\pi_2^\star = 0.4$ are dual optimal (we used KKT conditions)

Sensitivity interpretation of $\pi_1^\star$: if $b_1 = 1 + \epsilon$, optimal cost z increases by 0.4ϵ

Proof: For $b_1 = 1 + \epsilon$,
$x^\star = (0.9333 - 0.1333\epsilon, 0.1333 + 0.2667\epsilon, 0) \Rightarrow$ cost change equals $-1 \cdot 0.1333\epsilon + 2 \cdot 0.2667\epsilon = 0.4\epsilon$

Note: Expressing equality constraints as $-h(x) = 0$ gives $(-0.4, -0.4)$, note the change in sign of $\pi^\star$

Sign of Dual Multipliers

Dual optimal multiplier may be equal to

- sensitivity, or
- minus the sensitivity

of objective function $f_0(x)$ to change in right hand side of $f_i(x) \leq 0$

Sensitivity depends on how Lagrangian function is defined:

- If $L(x, \lambda) = f_0(x) - \lambda_i \cdot f_i(x)$ then then λ is equal to sensitivity
- If $L(x, \lambda) = f_0(x) + \lambda_i \cdot f_i(x)$ then λ equals minus sensitivity

Same idea applies for $h_i(x) = 0$