The gradient vector lives in the dual space

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms**x**ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: November 13, 2021 Last update: November 14, 2021

Optimization variable and gradient live in different spaces

- Consider the optimization problem

$$(\mathcal{P}) \ : \ \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x})$$

  - $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, where $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$
  - $g : \mathbb{R}^m \to \bar{\mathbb{R}}$
  - $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable
  - $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a real matrix (linear map / linear transformation)

- Assume $f \in C^1$ (f is differentiable), we can compute $\nabla f(\mathbf{x})$ for any $\mathbf{x}$

- A (non-trivial) fact: the variable $\mathbf{x}$ and the gradient $\nabla f(\mathbf{x})$ technically **do not live in the same space**. If $\mathbf{x}$ lives in the primal space, then $\nabla f(\mathbf{x})$ lives in the dual space.

- This document: explain what it is from optimization point of view.

Prerequisite of this document

- To understand this document, you need to
  - Fenchel conjugate
    The Fenchel conjugate of a function f, denoted as f^*, is defined as
    $$f^*(\mathbf{x}) = \sup_{\mathbf{u}} \langle \mathbf{u}, \mathbf{x} \rangle - f(\mathbf{u}).$$
  - Primal-Dual problem pair expressed in Fenchel conjugate
    $$(\mathcal{P}) \ : \ p = \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) \qquad \text{Primal problem}$$
    $$(\mathcal{D}) \ : \ d = \max_{\mathbf{y}} -f^*(\mathbf{A}^\top \mathbf{y}) - g^*(-\mathbf{y}) \qquad \text{Dual problem}$$
    Notation:
    - $p =$ optimal primal cost, $d =$ optimal dual cost
    - $\mathbf{x}^*$ optimal primal point, $\mathbf{y}^*$ optimal dual point
  - Condition (Sufficiency) for strong duality
    Under the following assumptions on f, g, A, we have $p = d$.
    1. f, g are convex function
    2. $\mathbf{A}\mathrm{dom} f \bigcap \mathrm{cont} g \neq \varnothing$[1]

[1] The intersection of domain of f, after a linear transformation by the matrix $\mathbf{A}$, with the set of $\mathbf{y}$ where g is finite and continuous, is non-empty.

A primal-dual optimality condition

- For the primal-dual problem

$$
\begin{aligned}
(\mathcal{P}) &: \; p = \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) \\
(\mathcal{D}) &: \; d = \max_{\mathbf{y}} -f^*(\mathbf{A}^\top \mathbf{y}) - g^*(-\mathbf{y})
\end{aligned}
$$

- Condition (Sufficiency) for strong duality
  1. f, g are convex function
  2. $\mathbf{A}\mathrm{dom} f \bigcap \text{cont} g \neq \varnothing$

- **Theorem** The points $\mathbf{x}^* \in \mathbb{R}^n, \mathbf{y}^* \in \mathbb{R}^m$ are the optimal points for $(\mathcal{P})$ and $(\mathcal{D})$, resp., if and only if they satisfy the following primal-dual optimality condition:

$$
\mathbf{A}^\top \mathbf{y}^* \in \partial f(\mathbf{x}^*) \quad \text{and} \quad -\mathbf{y}^* \in \partial g(\mathbf{A}\mathbf{x}^*).
$$

A closer look at the primal-dual optimality condition

- Now look at the primal-dual optimality condition:

$$\mathbf{A}^\top \mathbf{y}^* \in \partial f(\mathbf{x}^*) \quad \text{and} \quad -\mathbf{y}^* \in \partial g(\mathbf{A}\mathbf{x}^*).$$

- For simplicity, assume that f is differentiable. Then we have $\partial f(\mathbf{x}^*) = \{\nabla f(\mathbf{x}^*)\}$.

- For f is differentiable, the first part of the primal-dual optimality condition becomes

$$\mathbf{A}^\top \mathbf{y}^* = \nabla f(\mathbf{x}^*)$$

Why the gradient vector lives in the dual space

$$\mathbf{A}^\top \mathbf{y}^* = \nabla f(\mathbf{x}^*)$$

- The above equation means: after a linear transform (indicated by the matrix $\mathbf{A}$), the optimal dual vector ($\mathbf{y}^*$) equals to the gradient of f at the optimal point (at $\mathbf{x}^*$).
- Since we are equating the gradient ∇f with a dual vector $\mathbf{y}$ (who lives in the dual space), therefore, gradient is a dual vector.
- The linear transform $\mathbf{A}^\top$ does not change the fact that $\mathbf{A}^\top \mathbf{y}^*$ is a dual vector simply because $\mathbf{A}$ is just a linear transformation within the dual space.

Another explanation of why the gradient vector lives in the dual space

- Now we look at the fundamental definition of the gradient:

  Given a scalar function f and a inner product $\langle \cdot, \cdot \rangle$, the gradient of f is defined as the unique vector field whose dot product with any vector $\mathbf{v}$ at each point $\mathbf{x}$ is the directional derivative of f along $\mathbf{v}$. Mathematically

  $$\langle \nabla f(\mathbf{x}), \mathbf{v} \rangle = df(\mathbf{v}), \tag{$*$}$$

  where $df(\mathbf{v})$ is the total derivative of f along the curve fitting the vector $\mathbf{v}$.

- The important conceptual point here is that derivative is based on an inner product, which induces a metric.

- Equation $(*)$ simply means that gradient is defined as a linear functional (linear transformations / linear map) .

- By definition, a dual space $\mathcal{E}^*$ of a given primal space $\mathcal{E}$ is the space that consists of all linear transformations from $\mathcal{E}$ to $\mathcal{E}$.

- As gradient is a linear function in the primal space $\implies$ gradient is a dual vector.

Last page - summary

- Gradient is a dual vector explained in two ways.
- Next: it turns out, “gradient is a dual vector” is exactly one of the motivation of Mirror Descent, invented by Nemirovski and Yudin in 1983.
- Another motivation of mirror descent is about geometry, see here.

End of document