Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 4, Lecture 5: Generalized Inner Products

March 1, 2019 *University of Illinois at Urbana-Champaign*

1 Characterization of inner products

For our purposes, an inner product $\star$ is some function that maps two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ to a single value $\mathbf{x} \star \mathbf{y} \in \mathbb{R}$, satisfying the following axioms:

(a) Linearity in both arguments: for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$,

$$(a\mathbf{x} + b\mathbf{y}) \star \mathbf{z} = a(\mathbf{x} \star \mathbf{z}) + b(\mathbf{y} \star \mathbf{z}) \text{ and } \mathbf{x} \star (a\mathbf{y} + b\mathbf{z}) = a(\mathbf{x} \star \mathbf{y}) + b(\mathbf{x} \star \mathbf{z}).$$

(b) Symmetry: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x} \star \mathbf{y} = \mathbf{y} \star \mathbf{x}.$$

(c) Positivity: for all $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} \star \mathbf{x} \geq 0$$

with equality only if $\mathbf{x} = \mathbf{0}$.

(Some comments: first, if we have symmetry, then the second half of linearity follows from the first half, but I've included both halves in case we don't have symmetry. Second, linearity can be simplified to the two cases $(a\mathbf{x}) \star \mathbf{y} = a(\mathbf{x} \star \mathbf{y})$ and $(\mathbf{x} + \mathbf{y}) \star \mathbf{z} = (\mathbf{x} \star \mathbf{z}) + (\mathbf{y} \star \mathbf{z})$. Third, positivity doesn't need to say that $\mathbf{0} \star \mathbf{0} = 0$, since we can use linearity to show that $\mathbf{0} \star \mathbf{x} = \mathbf{x} \star \mathbf{0} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$.)

If we have an inner product $\star$, we can use it to define a norm $\|\cdot\|_\star$ by $\|\mathbf{x}\|_\star = \sqrt{\mathbf{x} \star \mathbf{x}}$. We could also write down the axioms that a norm has to satisfy, but it turns out that there are lots of possible norms satisfying those axioms, which are hard to describe.

On the other hand, it's easy to describe all inner products (and therefore all norms that come from inner products):

Theorem 1.1. *An operation $\star : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an inner product if and only if it can be written as*

$$\mathbf{x} \star \mathbf{y} = \mathbf{x}^\mathsf{T} H \mathbf{y}$$

for some positive definite $n \times n$ matrix H.

Proof. More precisely, the three properties of the inner product give us more and more information about H, so let's take them one at a time.

[1]This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-spring-2019.html

If we just have any "linear form"—if $\star$ satisfies condition (a), linearity—then we can put it in the form $\mathbf{x} \star \mathbf{y} = \mathbf{x}^\mathsf{T} H \mathbf{y}$ for some $n \times n$ matrix H. Conversely, by linearity of matrix multiplication, any expression of the form satisfies condition (a).

To see this, let $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \dots, \mathbf{e}^{(n)}$ be the n standard basis vectors of $\mathbb{R}^n$, and let H be the matrix given by $H_{ij} = \mathbf{e}^{(i)} \star \mathbf{e}^{(j)}$. Then we have

$$\begin{aligned}
\mathbf{x} \star \mathbf{y} &= \left(\sum_{i=1}^n x_i \mathbf{e}^{(i)}\right) \star \left(\sum_{j=1}^n y_j \mathbf{e}^{(j)}\right) \\
&= \sum_{i=1}^n x_i \left(\mathbf{e}^{(i)} \star \left(\sum_{j=1}^n y_j \mathbf{e}^{(j)}\right)\right) \\
&= \sum_{i=1}^n x_i \left(\sum_{j=1}^n y_j (\mathbf{e}^{(i)} \star \mathbf{e}^{(j)})\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i H_{ij} y_j = \mathbf{x}^\mathsf{T} H \mathbf{y}.
\end{aligned}$$

If $\star$ also satisfies condition (b), symmetry, then the matrix H we already have must be symmetric. Conversely, if H is symmetric, then the operation $\mathbf{x} \star \mathbf{y} = \mathbf{x}^\mathsf{T} H \mathbf{y}$ satisfies condition (b).

This follows from the way we defined the entries of H: $H_{ij} = \mathbf{e}^{(i)} \star \mathbf{e}^{(j)}$ while $H_{ji} = \mathbf{e}^{(j)} \star \mathbf{e}^{(i)}$. So if $\star$ is symmetric, then $H_{ij} = H_{ji}$, making $H = H^\mathsf{T}$. Conversely, if $H = H^\mathsf{T}$, then

$$\mathbf{x} \star \mathbf{y} = \mathbf{x}^\mathsf{T} H \mathbf{y} = \left(\mathbf{x}^\mathsf{T} H \mathbf{y}\right)^\mathsf{T} = \mathbf{y}^\mathsf{T} H^\mathsf{T} \mathbf{x} = \mathbf{y}^\mathsf{T} H \mathbf{x} = \mathbf{y} \star \mathbf{x}.$$

(We get to take the transpose of $\mathbf{x}^\mathsf{T} H \mathbf{y}$ for free because it's a 1×1 matrix, so it's always symmetric.)

Finally, $\star$ also satisfies condition (c), positivity, precisely when H is positive definite. In fact, the condition $\mathbf{x}^\mathsf{T} H \mathbf{x} \geq 0$ with equality only if $\mathbf{x} = \mathbf{0}$ is simply the definition of a positive definite matrix. $\square$

This takes some of the mystery out of inner products. In fact, we already know that a symmetric matrix H is positive definite if and only if we can write $H = B^\mathsf{T} B$ for some invertible matrix B. Then we have

$$\mathbf{x} \star \mathbf{y} = \mathbf{x}^\mathsf{T} H \mathbf{y} = \mathbf{x}^\mathsf{T} B^\mathsf{T} B \mathbf{y} = (B\mathbf{x})^\mathsf{T} B \mathbf{y} = (B\mathbf{x}) \cdot (B\mathbf{y}).$$

If we think of B as a change-of-basis matrix, this tells us that any inner product $\star$ is just the usual dot product $\cdot$, but taken in a different basis from the standard basis.

Nevertheless, we can still make good use of these generalized inner products, because the matrix B for a given positive definite matrix H is not easy to find, and doesnt always have a nice form when we do find it.

From now on, we'll use the notation $\mathbf{x} \cdot_H \mathbf{y}$ to write the inner product $\mathbf{x}^\mathsf{T} H \mathbf{y}$, and $\|\mathbf{x}\|_H$ for the associated norm

$$\|\mathbf{x}\|_H = \sqrt{\mathbf{x} \cdot_H \mathbf{x}} = \sqrt{\mathbf{x}^\mathsf{T} H \mathbf{x}}.$$

2 Applications

We can use generalized inner products to squeeze some more mileage out of our solutions to the optimization problems in this chapter.

For example, suppose we have the problem

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \mathbf{x}^\mathsf{T} H \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}. \end{aligned}$$

for a positive definite matrix H (and an $m \times n$ matrix A and vector $\mathbf{b} \in \mathbb{R}^m$). We now know that $\mathbf{x}^\mathsf{T} H\mathbf{x}$ is the generalized inner product $\mathbf{x} \cdot_H \mathbf{x} = \|\mathbf{x}\|_H^2$, and minimizing it is equivalent to minimizing the H-norm $\|\mathbf{x}\|_H$.

When $\|\mathbf{x}\|_H$ was the usual norm $\|\mathbf{x}\|$, the key to solving the problem was the perpendicularity condition: $\mathbf{x}^*$ is the minimum-norm solution if and only if $\mathbf{x}^* \cdot \mathbf{y} = 0$ for all $\mathbf{y} \in \mathbb{R}^n$ such that $A\mathbf{y} = \mathbf{0}$. We get a similar condition here for free:

Lemma 2.1. *A point $\mathbf{x}^*$ satisfying $A\mathbf{x}^* = \mathbf{b}$ is the minimum-H-norm solution to $A\mathbf{x} = \mathbf{b}$ if and only if*

$$\mathbf{x}^* \cdot_H \mathbf{y} = 0$$

for all $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{0}$.

We might imagine writing a proof of this lemma that repeats everything we've done in the past few lectures. In fact, as long as the only properties of $\cdot$ that we used in our previous proof are the properties (a), (b), and (c) that also hold for $\cdot_H$, and I promise you that that's the case, we don't need to: our previous proof already works here.

How does this translate into a method for solving the problem? Well, we have

$$\mathbf{y} \cdot_H \mathbf{x}^* = \mathbf{y}^\mathsf{T} H\mathbf{x}^* = \mathbf{y} \cdot (H\mathbf{x}^*)$$

so we want $H\mathbf{x}^*$ to be orthogonal to the null space $\{\mathbf{y} \in \mathbb{R}^n : A\mathbf{y} = \mathbf{0}\}$. Just as before, this means that $H\mathbf{x}^*$ has the form $A^\mathsf{T}\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^m$ or, in other words, $\mathbf{x}^*$ has the form $H^{-1}A^\mathsf{T}\mathbf{w}$.

Theorem 2.1. *The minimum-H-norm solution $\mathbf{x}^*$ of the underconstrained system $A\mathbf{x} = \mathbf{b}$ can be found by solving*

$$AH^{-1}A^\mathsf{T}\mathbf{w} = \mathbf{b}$$

for $\mathbf{w}$, then setting $\mathbf{x}^* = H^{-1}A^\mathsf{T}\mathbf{w}$.

We're phrasing this as a problem about some weird generalized norms, but in fact it lets us optimize many quadratic objective functions over the solution set to a linear equation.

2.1 Example problem

Suppose we want to minimize $3x^2 + 2xy + 2y^2$ subject to $3x - y = 3$. Okay, in this case, we can just set $y = 3x - 3$ and substitute and take derivatives, but that method rapidly becomes much more annoying when we have more equations to take into account.

The expression $3x^2 + 2xy + 2y^2$ is the square of the H-norm of (x, y) for the matrix $H = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. So the theorem tells us to solve

$$\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -1 \end{bmatrix} w = \begin{bmatrix} 3 \end{bmatrix}.$$

Simplifying, we get

$$\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -1 \end{bmatrix} w = \begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.6 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} w = \begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 1.4 \\ -1.2 \end{bmatrix} w = 5.4w$$

so we are just getting the equation $5.4w = 3$, or $w = \frac{5}{9}$.

So the optimal solution is $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1.4 \\ -1.2 \end{bmatrix} \frac{5}{9} = \begin{bmatrix} 7/9 \\ -2/3 \end{bmatrix}$.

3 More general cases

What about problems of the form

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad & \mathbf{x}^\mathsf{T} H \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}. \end{aligned}$$

where H is not positive definite? Here, the approach above isn't guaranteed to work, because $\cdot_H$ does not have the usual properties of an inner product. Such problems are outside the scope of this course, but here are a few possibilities.

- H may still turn out to be positive definite on the null space of A. (That is, for any $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{0}$, $\mathbf{y}^\mathsf{T} H \mathbf{y} \geq 0$, with equality only if $\mathbf{y} = \mathbf{0}$.) If so, then the theorem about minimum-H-norm solutions still applies.

- There may be some $\mathbf{y} \neq \mathbf{0}$ such that $A\mathbf{y} = \mathbf{0}$ but $\mathbf{y}^\mathsf{T} H \mathbf{y} < 0$.

  In this case, given any solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$, the point $\mathbf{x} + t\mathbf{y}$ is also a solution for any $t \in \mathbb{R}$, and $(\mathbf{x} + t\mathbf{y})^\mathsf{T} H (\mathbf{x} + t\mathbf{y})$ has a leading term of $(\mathbf{y}^\mathsf{T} H \mathbf{y}) t^2$ in t. This approaches $-\infty$ as $t \to \pm\infty$.

- H may be positive semidefinite on the null space of A, but there may be some $\mathbf{y} \neq \mathbf{0}$ such that $A\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^\mathsf{T} H \mathbf{y} = 0$.

  In this case, given any solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$, the point $\mathbf{x} + t\mathbf{y}$ is also a solution for any $t \in \mathbb{R}$, and $(\mathbf{x} + t\mathbf{y})^\mathsf{T} H (\mathbf{x} + t\mathbf{y})$ has a leading term of $(\mathbf{x} \cdot_H \mathbf{y}) t$. Sometimes this is guaranteed to be 0, in which case we proceed as usual; if it is nonzero, however, we can still get unbounded solutions as $t \to \infty$ or $t \to -\infty$.