Constrained optimization - theory

Øyvind Ryan

Apr 26, 2017

What is the gradient of $\|\boldsymbol{H}(\boldsymbol{x})\|^2$?

1. Write $\boldsymbol{G}(\boldsymbol{x}) = \|\boldsymbol{H}(\boldsymbol{x})\|^2 = \boldsymbol{F}(\boldsymbol{H}(\boldsymbol{x}))$, where $\boldsymbol{F}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = x_1^2 + ... + x_n^2$.
2. We have that $\boldsymbol{F}'(\boldsymbol{x}) = (2x_1, ..., 2x_n)$, so that $\boldsymbol{F}'(\boldsymbol{H}(\boldsymbol{x})) = 2\boldsymbol{H}(\boldsymbol{x})^T$.
3. The chain rule gives that $\boldsymbol{G}'(\boldsymbol{x}) = \boldsymbol{F}'(\boldsymbol{H}(\boldsymbol{x}))\boldsymbol{H}'(\boldsymbol{x}) = 2\boldsymbol{H}(\boldsymbol{x})^T\boldsymbol{H}'(\boldsymbol{x})$.
4. The Jacobi matrix of a real function is a row vector. The gradient is obtained by transposing this, so that $\nabla\boldsymbol{G}(\boldsymbol{x}) = 2\boldsymbol{H}'(\boldsymbol{x})^T\boldsymbol{H}(\boldsymbol{x})$.

Sketch of the proof of the non-negativity of μ in Theorem 5.3

1. Define the tangent vector space, and show that all elements $\boldsymbol{d}$ therein satsify $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$.
2. Define linearized feasible directions
3. Show that the tangent vector space, and the linearized feasible directions are the same when $\boldsymbol{x}^*$ is regular
4. Apply Farkas lemma.

The tangent vector cone

Let $C \subseteq \mathbb{R}^n$ and let $\boldsymbol{x} \in C$. A vector $\boldsymbol{d} \in \mathbb{R}^n$ is called a *tangent* (vector) to C at $\boldsymbol{x}$ if there is a sequence $\{\boldsymbol{x}^k\}$ in C and a sequence $\{\alpha_k\}$ in $\mathbb{R}_+$ such that

$$\lim_{k \to \infty} (\boldsymbol{x}^k - \boldsymbol{x})/\alpha_k = \boldsymbol{d}.$$

The set of tangent vectors at $\boldsymbol{x}$ is denoted by $T_C(\boldsymbol{x})$.

1. Let C be the set of feasible solutions (those $\boldsymbol{x}$ satisfying all the equality and inequality constraints).
2. One shows that $\boldsymbol{x}^*$ satisfies $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$ for all $\boldsymbol{d} \in T_C(\boldsymbol{x}^*)$.

Linearized feasible directions

A *linearized feasible direction* at $\boldsymbol{x} \in C$ is a vector $\boldsymbol{d}$ such that

$$\begin{aligned} \boldsymbol{d} \cdot \nabla h_i(\boldsymbol{x}) &= 0 \quad (i \leq m) \\ \boldsymbol{d} \cdot \nabla g_j(\boldsymbol{x}) &\leq 0 \quad (j \in A(\boldsymbol{x})). \end{aligned}$$

1. We denote by $LF_C(\boldsymbol{x})$ the set of all linearized feasible directions at $\boldsymbol{x}$.
2. since $\boldsymbol{H}'(\boldsymbol{x}^*)$ is the matrix with rows $\nabla h_i(\boldsymbol{x}^*)$, the first condition is the same as $\boldsymbol{H}'(\boldsymbol{x}^*)\boldsymbol{d} = 0$.
3. when all constraints are active the second condition is the same as $\boldsymbol{G}'(\boldsymbol{x}^*)\boldsymbol{d} \leq 0$.

Connection between tangent space and the feasible directions

Let $\boldsymbol{x}^* \in C$. Then $T_C(\boldsymbol{x}^*) \subseteq LF_C(\boldsymbol{x}^*)$. If $\boldsymbol{x}^*$ is a regular point, then $T_C(\boldsymbol{x}^*) = LF_C(\boldsymbol{x}^*)$.

Putting these things together, when $\boldsymbol{x}^*$ is regular, $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$ for all $\boldsymbol{d} \in LF_C(\boldsymbol{x}^*)$.

Farkas lemma

If B and C are matrices with n rows, and K is the cone defined by $K = \{B\boldsymbol{y} + C\boldsymbol{w}, \text{ with } \boldsymbol{y} \geq \boldsymbol{0}\}$, then exactly one of the following two alternatives are true:

1. $g \in K$
2. There exists a $\boldsymbol{d} \in \mathbb{R}^n$ so that $\boldsymbol{g}^T\boldsymbol{d} < 0$, $B^T\boldsymbol{d} \geq 0$, and $C^T\boldsymbol{d} = 0$.

Now, do the following:

1. Set $\boldsymbol{g} = \nabla f(\boldsymbol{x}^*)$, $B = -\boldsymbol{G}'(\boldsymbol{x}^*)^T$, and $C = -\boldsymbol{H}'(\boldsymbol{x}^*)^T$,
2. $B^T\boldsymbol{d} \geq 0$, and $C^T\boldsymbol{d} = 0$ simply says that $\boldsymbol{d} \in LF_C(\boldsymbol{x}^*) = T_C(\boldsymbol{x}^*)$.
3. For all $\boldsymbol{d} \in T_C(\boldsymbol{x}^*)$ we have proved that $\boldsymbol{g}^T\boldsymbol{d} = \nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} \geq 0$, so that point 2 of Farkas lemma does not hold for $\boldsymbol{g} = \nabla f(\boldsymbol{x}^*)$.
4. We conclude that $\boldsymbol{g} = \nabla f(\boldsymbol{x}^*) \in K$.

Conclusion

This means that we can find $\boldsymbol{y} \geq 0$ and $\boldsymbol{w}$ so that

$$\boldsymbol{g} = \nabla f(\boldsymbol{x}^*) = -\boldsymbol{H}'(\boldsymbol{x}^*)^T \boldsymbol{w} - \boldsymbol{G}'(\boldsymbol{x}^*)^T \boldsymbol{y} = B\boldsymbol{y} + C\boldsymbol{w}.$$

But this states exactly what we want to prove:

1. that $\nabla f(\boldsymbol{x}^*) + \boldsymbol{H}'(\boldsymbol{x}^*)^T \boldsymbol{w} + \boldsymbol{G}'(\boldsymbol{x}^*)^T \boldsymbol{y} = \boldsymbol{0}$,
2. that $\boldsymbol{w}$ contains the Lagrange multipliers λ_i,
3. that $\boldsymbol{y}$ contains the μ_i, and that they must be non-negative.