Constrained Optimization Theory

Stephen J. Wright[1]

[2]Computer Sciences Department,
University of Wisconsin-Madison.

IMA, August 2016

Constrained Optimization Theory

How can we recognize solutions of constrained optimization problems? Answering this question is crucial to algorithm design.

We already covered some cases, associated with the geometric formulation

$$\min f(x) \;\; \text{s.t.} \;\; x \in \Omega,$$

where Ω is closed and convex. We'll review these.

But we focus mainly on the case in which the feasible set is specified algebraically, that is,

$$\begin{aligned} c_i(x) &= 0, \quad \text{for all } i \in \mathcal{E}, \\ c_i(x) &\leq 0, \quad \text{for all } i \in \mathcal{I}. \end{aligned}$$

This general case admits a few complications! But ultimately we get a set of "checkable" conditions.

Closed Convex Ω

Recall: When Ω is closed and convex, we have the normal cone defined for all $x \in \Omega$ by

$$N_\Omega(x) := \{d \mid d^T(y-x) \le 0 \text{ for all } y \in \Omega\}.$$

Theorem

Suppose Ω is closed and convex and that f is convex with continuous gradients. Then a necessary and sufficient condition for x^* to be a (global) solution of $\min f(x)$ s.t. $x \in \Omega$ is that

$$-\nabla f(x^*) \in N_\Omega(x^*).$$

Proof.

Suppose that $-\nabla f(x^*) \in N_\Omega(x^*)$. By convexity we have for any $y \in \Omega$ that

$$f(y) \geq f(x^*) + \nabla f(x^*)^T (y - x^*) \geq f(x^*),$$

where the last condition arises from $-\nabla f(x^*) \in N_\Omega(x^*)$ and the definition of $N_\Omega(x^*)$. Thus x^* is a global solution.

Now suppose that $-\nabla f(x^*) \notin N_\Omega(x^*)$. Then there exists $y \in \Omega$ such that $-\nabla f(x^*)^T (y - x^*) > 0$. From Taylor's theorem and continuity of ∇f, we have for small $\alpha > 0$ that $x^* + \alpha(y - x^*) \in \Omega$ and

$$f(x^* + \alpha(y - x^*)) = f(x^*) + \alpha \nabla f(x^*)^T (y - x^*) + o(\alpha) < f(x^*),$$

where the latter holds for sufficiently small $\alpha > 0$. Thus, x^* is not a minimizer. □

Ω closed and convex, f smooth

When f is not convex, the condition $-\nabla f(x^*) \in N_\Omega(x^*)$ is still a necessary condition.

Theorem

Suppose Ω is closed and convex and that f has continuous gradients. Then if x^* is a local solution of $\min f(x)$ s.t. $x \in \Omega$, we have

$$-\nabla f(x^*) \in N_\Omega(x^*).$$

Proof.

Suppose that $-\nabla f(x^*) \notin N_\Omega(x^*)$. Then we use the second path of the previous proof to identify $y \in \Omega$ such that $f(x^* + \alpha(y - x^*)) < f(x^*)$ for all α sufficiently small and positive. *Any* neighborhood of x^* will contain a point of the form $x^* + \alpha(y - x^*)$ for some small enough $\alpha > 0$, and this point has a lower function value than $f(x^*)$. Thus, x^* is not a local solution. □

Tangent Cone

When Ω is specified algebraically and / or is nonconvex, we have the issue of how to define the normal cone $N_\Omega(x)$. We propose a more general definition (which coincides with the one above when Ω is convex).

The new definition is based on the Tangent cone $T_\Omega(x)$, which is defined as the cone of limiting feasible directions.

We say that d is a *limiting feasible direction* to Ω at x if there exists a sequence $\{z_k\}$ with $z_k \in \Omega$ for all k, and a sequence $\{t_k\}$ of positive scalars, such that

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d.$$

Do some nonconvex examples:

$$\Omega = \{(x_1, x_2) \,|\, x_2 \le x_1^2\}, \quad \Omega = \{(x_1, x_2) \,|\, -x_1^2 \le x_2 \le x_1^2\}.$$

Tangent Cone of a Polyhedron

When Ω is defined by linear constraints, it's easy to identify the tangent cone algebraically.

$$\Omega = \{x \,|\, a_i^T x = b_i, \; i \in \mathcal{E}; \;\; a_i^T x \le b_i, \; i \in \mathcal{I}\}.$$

At a given point $\bar{x} \in \Omega$, we define the active set:

$$\mathcal{A}(\bar{x}) := \mathcal{E} \cup \{i \in \mathcal{I} \,|\, a_i^T \bar{x} = b_i\}.$$

The tangent cone at $\bar{x}$ is then

$$T_\Omega(\bar{x}) = \{d \,|\, a_i^T d = 0, \; i \in \mathcal{E}; \;\; a_i^T d \le 0, \; i \in \mathcal{A}(\bar{x}) \cap \mathcal{I}\}.$$

Proof: For $i \in \mathcal{A}(\bar{x})$, we have

$$a_i^T d = \lim_{k \to \infty} \frac{a_i^T z_k - a_i^T \bar{x}}{t_k} = \lim_{k \to \infty} \frac{a_i^T z_k - b_i}{t_k}.$$

For $i \in \mathcal{E}$, we have $a_i^T z_k - b_i = 0$ for all k, so $a_i^T d = 0$. For $i \in \mathcal{I} \cap \mathcal{A}(\bar{x})$, we have $a_i^T z_k - b_i \le 0$, so in the limit we have $a_i^T d \le 0$.

Nonlinear Algebraic Constraints

What about nonlinear algebraic constraints

$$c_i(x) = 0, \; i \in \mathcal{E}; \quad c_i(x) \le 0, \; i \in \mathcal{I}.$$

Can we linearize these constraints at a given feasible $\bar{x}$, then use the polyhedron methodology to find $T_\Omega(\bar{x})$?

Following the previous slide, we can define the active set:

$$\mathcal{A}(\bar{x}) := \mathcal{E} \cup \{i \in \mathcal{I} \,|\, c_i(\bar{x}) = 0\}$$

and the feasible direction set:

$$\mathcal{F}(\bar{x}) := \{d \,|\, \nabla c_i(\bar{x})^T d = 0, \; i \in \mathcal{E}; \;\; \nabla c_i(\bar{x})^T d \le 0, \; i \in \mathcal{A}(\bar{x}) \cap \mathcal{I}\}.$$

But we don't necessarily have $\mathcal{F}(\bar{x}) = T_\Omega(\bar{x})$! We need extra conditions called constraint qualifications to guarantee that this equality holds.

Nonlinear Constraints

It's easy to prove that $T_\Omega(\bar{x}) \subset \mathcal{F}(\bar{x})$, provided the constraints c_i are smooth. This is a consequence of Taylor's theorem. (Exercise!)

The issue comes with proving the converse: $\mathcal{F}(\bar{x}) \subset T_\Omega(\bar{x})$. Here's where we need the constraint qualifications.

$$\text{Example:} \quad x_2 \le 0, \; x_2 \ge x_1^2.$$

Here we have a single feasible point: $x^* = (0,0)$. But

$$T_\Omega(0) = \{0\}, \quad \mathcal{F}(0) = \{(d_1, 0) \,|\, d_1 \in \mathbb{R}\}.$$

An even more elementary example: $x^3 \ge 0$. Here we have

$$T_\Omega(0) = \{d \,|\, d \ge 0\}, \quad \mathcal{F}(0) = \mathbb{R}.$$

Constraint Qualifications

Two famous CQs, that ensure that $\mathcal{F}(\bar{x}) \subset T_\Omega(\bar{x})$ are:

LICQ: (Linear Independent Constraint Qualification): the active constraint gradients $\nabla c_i(\bar{x})$, $i \in \mathcal{A}(\bar{x})$, are linearly independent. Can use the implicit function theorem to prove $\mathcal{F}(\bar{x}) \subset T_\Omega(\bar{x})$.

MFCQ: (Mangasarian-Fromovitz Constraint Qualification): The equality constraint gradients are linearly independent, and there exists a vector d such that

$$\nabla c_i(\bar{x})^T d = 0, \ \ i \in \mathcal{E}; \quad \nabla c_i(\bar{x})^T d < 0, \ \ i \in \mathcal{A}(\bar{x}) \cap \mathcal{I}.$$

(Note: Strict inequality!)

First-Order Necessary Condition

A fundamental necessary condition for x^* to be a local min is that

$$\nabla f(x^*)^T d \geq 0 \quad \text{for all } d \in T_\Omega(x^*).$$

(Prove using Taylor's theorem: If this condition does not hold, we can construct a sequence $\{z_k\}$ with $z_k \in \Omega$ and $f(z_k) < f(x^*)$, so x^* cannot be a local min.)

But as written, this condition is hard to check!

We now give the general definition of the normal cone:

$$N_\Omega(\bar{x}) = \{v \,|\, v^T d \leq 0 \ \text{ for all } d \in T_\Omega(\bar{x})\}.$$

That is, the normal cone is the polar of the tangent cone.

(Exercise: Show that this coincides with the earlier def for convex Ω!)

The first-order necessary condition then becomes $-\nabla f(x^*) \in N_\Omega(x^*)$!

CQs and First-Order Necessary

If a CQ holds, we have $\mathcal{F}(x^*) = T_\Omega(x^*)$. To calculate the normal cone in this case, we need to find the normal to the polyhedral set $\mathcal{F}(x^*)$ at x^*.

Here's where Farkas's Lemma is useful!

Farkas's Lemma: Given vectors a_i, $i = 1, 2, \dots, m$ and b, EITHER there are coefficient $\lambda_i \geq 0$ such that $b = \sum_{i=1}^m \lambda_i a_i$ OR there is a vector w such that $b^T w > 0$ and $a_i^T w \leq 0$, $i = 1, 2, \dots, m$.

We apply this Lemma defining the a_i vectors to be

$$\pm \nabla c_i(x^*),\ i \in \mathcal{E}; \quad \nabla c_i(x^*),\ i \in \mathcal{A}(x^*) \cap \mathcal{I},$$

and $b = -\nabla f(x^*)$.

KKT Conditions

The first order necessary condition tells us that there is no vector d with $-\nabla f(x^*)^T d > 0$ and

$$d^T \nabla c_i(x^*) = 0, \; i \in \mathcal{E}; \quad d^T \nabla c_i(x^*) \le 0, \; i \in \mathcal{A}(x^*) \cap \mathcal{I}.$$

Farkas's Lemma tells us that the alternative must be true, that is, there are nonnegative coefficients such that

$$-\nabla f(x^*) = \sum_{i \in \mathcal{E}} (\lambda_i^+ - \lambda_i^-) \nabla c_i(x^*) + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i \nabla c_i(x^*).$$

By combining $\lambda_i := \lambda_i^+ - \lambda_i^-$ for $i \in \mathcal{E}$, we obtain the condition:

$$-\nabla f(x^*) = \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(x^*) + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i \nabla c_i(x^*).$$

We combine this equality with other conditions that ensure feasibility of x^* to obtain the KKT conditions.

KKT Conditions: First-Order Necessary

Theorem

Suppose that x^* is a local solution of the problem

$$\min f(x) \ \ s.t. \ \ c_i(x) = 0, \ i \in \mathcal{E}, \ \ c_i(x) \le 0, \ i \in \mathcal{I},$$

and that a constraint qualification (linear constraints, LICQ, MFCQ) is satisfied at x^*. Then there are coefficients λ_i, $i \in \mathcal{E} \cup \mathcal{I}$ such that the following are true:

$$-\nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(x^*),$$

$$0 \le \lambda_i \perp c_i(x^*) \le 0, \quad i \in \mathcal{I},$$

$$c_i(x^*) = 0, \quad i \in \mathcal{E}.$$

Note that the $\perp$ condition forces $\lambda_i = 0$ for $i \notin \mathcal{A}(x^*)$.

Lagrangian

The Lagrangian is a linear combination of objective function and constraints, with coefficients λ_i called Lagrange multipliers.

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x).$$

We can restate the first of the KKT conditions succinctly using the Lagrangian:

$$\nabla_x \mathcal{L}(x^*, \lambda) = 0.$$

It's also useful in defining second-order conditions that characterize solutions.

Unconstrained: Second-order Conditions

Go back to the unconstrained problem min $f(x)$ with f smooth.

If f is convex, the first-order condition $\nabla f(x^*) = 0$ is necesssary and sufficient for x to be a solution of the problem.

If f is nonconvex, we can't say much about global solutions (except in special cases), but we can talk about conditions for local solutions.

We saw earlier than $\nabla f(x^*) = 0$ is a first-order necessary (1oN) condition for a local solution. We can also identify second-order conditions that make reference to the Hessian $\nabla^2 f(x^*)$.

Second-Order Necessary (2oN) Conditions

Theorem

If x^* is a local minimizer of f and $\nabla^2 f$ is continuous in an open neighborhood of x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite.

Proof.

We know from 1oN condition that $\nabla f(x^*) = 0$ Assume for contradiction that there is a direction p such that $p^T \nabla^2 f(x^*) p < 0$. By Taylor's theorem, we have that

$$f(x^* + \alpha p) = f(x^*) + \frac{1}{2}\alpha^2 p^T \nabla^2 f(x^*) p + o(\alpha^2) < f(x^*)$$

for all $\alpha > 0$ sufficiently small. Thus x^* cannot be a local solution. □

Second-Order Sufficient (2oS) Conditions

Theorem

Suppose that $\nabla^2 f$ is continuous in a neighborhood of x^* and that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite. Then x^* is a strict local minimizer of f.

The proof follows again from Taylor's theorem, using a second-order expansion around x^*.

Critical Cone

An important quantity in the second-order conditions is the critical cone $\mathcal{C}(x^*, \lambda^*)$. This the cone of directions w in the linearized feasible set $\mathcal{F}(x^*)$ for which the KKT conditions alone do not tell us whether f increases along w. We need higher-order information to resolve the issue.

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) \,|\, \nabla c_i(x^*)^T w = 0, \;\; i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ for which } \lambda_i^* > 0\}$$

This excludes directions $w \in \mathcal{F}(x^*)$ such that $\nabla c_i(x^*)^T w < 0$ for some $\lambda_i^* > 0$, $i \in \mathcal{I}$.

For $w \in \mathcal{C}(x^*, \lambda^*)$ we have from KKT that

$$w^T \nabla f(x^*) = - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(x^*)^T w = 0,$$

so the first-order conditions alone are not enough to verify that w is an ascent direction for f. (Other directions in $\mathcal{F}(x^*)$ yield $w^T \nabla f(x^*) > 0$.)

Second-Order Necessary (2oN) Conditions

Theorem

Suppose that x^* is a local solution at which a CQ holds, and suppose that KKT conditions are satisfied by (x^*, λ^*). Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*).$$

Proofs uses the fact that w is a limiting feasible direction, Taylor's theorem applied to $\mathcal{L}(\cdot, \lambda^*)$, definition of $\mathcal{C}(x^*, \lambda^*)$. See (Nocedal and Wright, 2006, Theorem 12.5).

Second-Order Sufficient (2oS) Conditions

Theorem

Suppose that x^* is feasible and there exists λ^* such that (x^*, λ^*) satisfy KKT. Suppose that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0 \quad \textit{for all } w \in \mathcal{C}(x^*, \lambda^*) \textit{ with } w \neq 0.$$

Then x^* is a strict local solution.

Proof uses Taylor's theorem, compactness of $\{d \in \mathcal{C}(x^*, \lambda^*) \,|\, \|d\| = 1\}$.

Note the differences between 2oN and 2oS results:

- strict inequality in the curvature condition;
- strict local solution in the 2oS result, not just local solution;
- 2oS result does not require a CQ.

References I

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York.