CMO: Constrained Optimization

Eklavya Sharma

In constrained optimization, we have to find

$$x^* = \operatorname*{argmin}_{x \in C} f(x)$$

where $C \in \mathbb{R}^d$ is a closed set. C is called the feasible region. We say that x is feasible iff $x \in C$.

The methods which we developed for unconstrained optimization often don't work for constrained optimization because properties of optimal solutions are different here. For example, if x^* is an unconstrained minimum of f, then $\nabla_f(x^*) = 0$. This doesn't hold for constrained minima. $\min_{x \in [1,2]} x^2$ is an example.

We'll consider several special cases of constrained optimization.

Contents

1 Introduction

Definition 1 (Feasible directions)**.** *$u \in \mathbb{R}^d$ is feasible direction at $x \in C$ iff*

$$\exists \overline{\alpha} > 0, \forall \alpha \in [0, \overline{\alpha}], x + \alpha u \in C$$

The set of feasible directions at x is denoted by $\mathrm{FS}(x)$.

Theorem 1. *If x is a local minimum of f, then there is no feasible descent direction. Formally,*

$$\forall u \in \mathrm{FS}(x), \nabla_f(x)^T u \geq 0$$

Proof Sketch. If there is a feasible descent direction u at x, then for any arbitrarily small α, we can decrease f by moving α distance towards u. So f is not a local minimum. □

Note that the converse need not be true. Let x be a saddle point of f and let there be no constraints. Then every direction is not a descent direction (and not an ascent direction) but x is not a local minimum.

2 Projection onto a convex set

Theorem 2. *Let C be a convex set. Let $x^* = \operatorname{argmin}_{x \in C} f(x)$. Then*

$$\forall x \in C, x - x^* \in \mathrm{FS}(x^*)$$

In this section, we'll now fix the objective function to be $f(x) = \frac{1}{2}\|x - z\|^2$ and consider the feasible region C to be convex. Also, assume that $z \notin C$.

Definition 2 (Projection)**.** *Let $x^* = \operatorname{argmin}_{x \in C} f(x)$. Then x^* is called the projection of z onto C.*

Theorem 3.

$$x^* = \operatorname*{argmin}_{x \in C} f(x) \iff (\forall x \in C, (x^* - z)^T(x - x^*) \geq 0)$$

Proof. Let $x^* = \operatorname{argmin}_{x \in C}$. By theorem 2, we get that

$$\forall x \in C, x - x^* \in \mathrm{FS}(x^*)$$

By theorem 1, we get that

$$\begin{aligned} &\forall x \in C, \nabla_f(x^*)^T(x - x^*) \geq 0 \\ &\implies \forall x \in C, (x^* - z)^T(x - x^*) \geq 0 \end{aligned}$$

Now assume that $\forall x \in C, (x^* - z)^T(x - x^*) \geq 0$.

$$\begin{aligned} f(x) &= \frac{1}{2}\|(x - x^*) + (x^* - z)\|^2 \\ &= \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|x^* - z\|^2 + (x^* - z)^T(x - x^*) \\ &\geq f(x^*) \end{aligned}$$

Therefore, $x^* = \operatorname{argmin}_{x \in C} f(x)$. □

Theorem 4. *There is a half-space which separates C and z. Formally,*

$$\forall x \in C, w^T x > w^T z$$

where $w = x^* - z$.

Proof.

$$\begin{aligned} &(x^* - z)^T(x - x^*) \geq 0 && \text{(by theorem 3)} \\ &\implies (x^* - z)^T x \geq (x^* - z)^T x^* \\ &\qquad \geq (x^* - z)^T(x^* - z + z) \\ &\qquad \geq \|x^* - z\|^2 + (x^* - z)^T z \\ &\qquad > (x^* - z)^T z \end{aligned}$$

□

3 Inequality constraints

Define the feasible region as

$$C = \{x : (\forall i \in I, c_i(x) \geq 0) \wedge (\forall i \in E, h_i(x) = 0)$$

Here $\{c_i : i \in I\}$ is the set of inequality constraints and $\{h_i : i \in I\}$ is the set of equality constraints. Since we can write the constraint $h_i(x) = 0$ as the 2 constraints $h_i(x) \geq 0$ and $-h_i(x) \geq 0$, we'll ignore equality constraints for now.

Our minimization algorithm will iteratively choose a feasible descent direction and make a small step in that direction.

By the definition of feasible direction, we get

$$u \in \mathrm{FS}(x) \iff \exists \overline{\alpha} > 0, \forall \alpha \in [0, \overline{\alpha}], c_i(x + \alpha u) \geq 0$$

Also, for $x \in C$, define LFS (called linearized feasible directions) as

$$\mathrm{LFS}(x) = \bigcap_{i \in I} \begin{cases} \mathbb{R}^d & \text{if } c_i(x) > 0 \\ \{u : \nabla_{c_i}(x)^T u \geq 0\} & \text{if } c_i(x) = 0 \end{cases}$$

Intuitively, LFS should be the same as FS. Unfortunately, they need not be the same.

Define descent directions (DS) as

$$u \in \mathrm{DS}(x) \iff \nabla_f(x)^T u < 0$$

When $\mathrm{FS}(x) \cap \mathrm{DS}(x) = \mathrm{LFS}(x) \cap \mathrm{DS}(x)$, we say that x is regular. Regularity always holds when the constraints are linear.

At a point x, a constraint c_i is said to be active iff $c_i(x) = 0$.

Theorem 5 (Farkas' Lemma)**.** *Let A be a d by m matrix and $b \in \mathbb{R}^d$. For a vector x, let $x \geq 0$ mean that all components of x are non-negative. Let $T = \{u \mid b^T u < 0 \wedge A^T u \geq 0\}$. Let $L = \{\lambda \mid b = A\lambda \wedge \lambda \geq 0\}$. Then $T = \{\} \iff L \neq \{\}$.*

Let I' be the set of active constraints at x^*. Let $|I'| = m$. Let A be the matrix whose i^{th} column is $\nabla_{c_i}(x^*)$. Then A is a d by m matrix. Let $b = \nabla_f(x^*)$. Then

$$u \in \mathrm{LFS}(x^*) \iff A^T u \geq 0 \qquad\qquad u \in \mathrm{DS}(x^*) \iff b^T u < 0$$

Then by Farkas' lemma, we get that

$$\mathrm{LFS}(x^*) \cap \mathrm{DS}(x^*) = \{\} \iff (\exists \lambda \geq 0, b = A\lambda)$$

For such a λ, we have

$$\nabla_f(x^*) = A\lambda = \sum_{i \in I'} \lambda_i \, \nabla_{c_i}(x^*)$$

This is equivalent to saying that

$$\nabla_f(x^*) = \sum_{i \in I} \lambda_i \, \nabla_{c_i}(x^*) \quad \text{ where } \lambda_i c_i(x^*) = 0$$

If x^* is a local minimum and a regular point, then $\mathrm{LFS}(x^*) \cap \mathrm{DS}(x^*) = \{\}$. So there exists $\lambda \in \mathbb{R}^m$ such that

- (Primal feasibility) $\forall i \in I, c_i(x^*) \geq 0$.
- (Stationarity) $\nabla_f(x^*) = \sum_{i \in I} \lambda_i \nabla_{c_i}(x^*)$.
- (Dual feasibility) $\forall i \in I, \lambda_i \geq 0$.
- (Complementary slackness) $\forall i \in I, \lambda_i c_i(x^*) = 0$.

These 4 conditions are called 'KKT conditions'. When these conditions hold for x and λ, (x, λ) is said to be a KKT point.

This is generally stated using the Lagrangian function (we're also going to consider the equality constraints now):

$$L(x, \lambda, \mu) = f(x) - \lambda^T c(x) - \mu^T h(x)$$

- (Primal feasibility) $c(x^*) \geq 0$ and $h(x^*) = 0$.
- (Stationarity) $\nabla_x L(x, \lambda, \mu) = 0$.
- (Dual feasibility) $\lambda \geq 0$.
- (Complementary slackness) $\forall i \in I, \lambda_i c_i(x^*) = 0$.