Constrained Optimization

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

March 18, 2024

Motivation

Manufacturing

- Suppose we have m different materials; we have s_i units of each material i in stock.
- We can manufacture k different products; product j gives us profit p_j and uses c_{ij} amount of material i to make.
- To maximize profits, we can solve the following optimization problem for the total amount x_j we should manufacture of each item j:

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} \sum_{j=1}^{k} p_j x_j \\ \text{such that } & x_j \geq 0 \ \forall\ j \in \{1, 2, \ldots, k\} \\ & \sum_{j=1}^{k} c_{ij} x_j \leq s_i, \ \forall\ i \in \{1, 2, \ldots, m\} \end{aligned} \tag{1}$$

- The first constraint ensures that we do not make negative numbers of any product,
- and the second ensures that we do not use more than our stock of each material.

Constrained Problem

A general formulation of these problems is:

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{2}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices.

- x is a vector of variables, also called unknown or parameters;
- f is the objective function, a function of x that we want to optimise (minimise or maximise);
- c is the vector function of constraints that must be satisfied by the unknowns x.
- c_i, $i \in \mathscr{E}$ are the equality constraints.
- c_i, $i \in \mathscr{I}$ are the inequality constraints.

Compact form of Constrained Problem

Definition

Define the feasible set Ω to be the set of points x that satisfy the constraints; that is,

$$\Omega = \{x \mid c_i(x) = 0, \quad i \in \mathcal{E}; \quad c_i(x) \geq 0, \quad i \in \mathcal{I}\}, \tag{3}$$

Now (2) can be rewritten more compactly as:

Constrained Problem

$$\min_{x \in \Omega} f(x). \tag{4}$$

Characterizations of the Solutions

- For the unconstrained optimization problems the solution point x^* was characterised in the following way:
- Necessary conditions: Local minima of unconstrained problems have

$$\nabla f(x^*) = 0$$

  and,

$$\nabla^2 f(x^*) \quad \text{is positive semidefinite}$$

- Sufficient conditions: Any point x^* at which $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite is a strong local minimiser of f.

LOCAL AND GLOBAL SOLUTIONS

- We have seen already that global solutions are difficult to find even when there are no constraints.
- The situation may improve when we add constraints.
- The feasible set might exclude many of the local minima.
- It might be comparatively easy to pick the global minimum from those that remain.

LOCAL AND GLOBAL SOLUTIONS

- Consider the problem

$$\min_{x \in \mathbb{R}^n} ||x||_2^2, \quad \text{subject to } ||x||_2^2 \geq 1. \tag{5}$$

- Without the constraint, this is a convex quadratic problem with unique minimiser $x = 0$.
- When the constraint is added, any vector x with $||x|| = 1$ solves the problem.
- There are infinitely many such vectors (hence, infinitely many local minima) whenever $n \geq 2$

LOCAL AND GLOBAL SOLUTIONS

- Addition of a constraint produces a large number of local solutions that do not form a connected set.
- Consider

$$\min_{x \in \mathbb{R}^2} (x_2 + 100)^2 + 0.01 x_1^2, \quad \text{subject to } x_2 - \cos x_1 \geq 0, \tag{6}$$

- Without the constraint, the problem has the unique solution $(-100, 0)$.
- With the constraint there are local solutions near the points

$$(x_1, x_2) = (k\pi, -1), \quad \text{for } k = \pm 1, \pm 3, \pm 5, \ldots$$

LOCAL AND GLOBAL SOLUTIONS

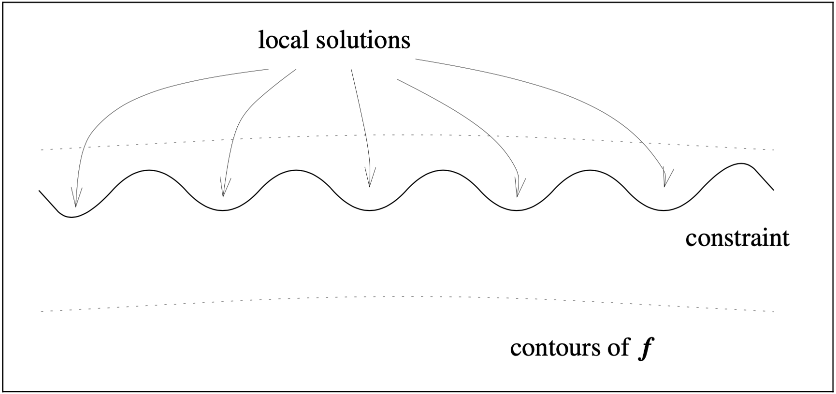

Figure Constrained problem with many isolated local solutions.

LOCAL AND GLOBAL SOLUTIONS

- Local and global solutions are defined in a very similar fashion as they were for the unconstrained case.
- The new caveat that comes into action in the definitions for the constrained case is the inclusion of constraints leading to a restriction imposed via a feasible set (space).

Definition

A vector x^* is a local solution of the constrained minimisation problem (4) if $x^* \in \Omega$ and there exists a neighbourhood $\mathscr{N}$ of x^* such that

$$f(x^*) \leq f(x) \quad \text{for all } x \in \Omega \cap \mathscr{N}$$

LOCAL AND GLOBAL SOLUTIONS

Definition

A vector x^* is called a strict local solution (also called a strong local solution) if $x^* \in \Omega$ and there is a neighbourhood $\mathcal{N}$ of x^* such that

$$f(x^*) < f(x) \quad \text{for all } x \in \mathcal{N} \cap \Omega \quad \text{with } x \neq x^*$$

Definition

A point x^* is an isolated local solution if $x^* \in \Omega$ and there is a neighbourhood $\mathcal{N}$ of x^* such that x^* is the only local minimiser in $N \cap \Omega$.

Smoothness

- Smoothness of objective functions and constraints is an important issue in characterizing solutions.
- Just as in the unconstrained case, it ensures that the objective function and the constraints all behave in a reasonably predictable way.
- Allows algorithms to make good choices for search directions.
- Non-smooth functions contain “kinks” or “jumps” where the smoothness breaks down.
- The feasible region for any given constrained optimization problem usually contains many kinks and sharp edges.

Smoothness

- Does this mean that the constraint functions that describe these regions are non-smooth?

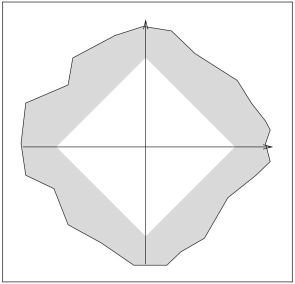

Figure: A feasible region with a non-smooth boundary can be described by smooth constraints.

- The answer is often no, because the non-smooth boundaries can often be described by a collection of smooth constraint functions.

Smoothness

- The figure above shows a diamond-shaped feasible region in $\mathbb{R}^2$.
- It could be described by the single non-smooth constraint

$$||x||_1 = |x_1| + |x_2| \leq 1.$$

- Or, it could also be brought out as an intersection of four smooth (in fact, linear) constraints:

$$x_1 + x_2 \leq 1, \quad x_1 - x_2 \leq 1, \quad -x_1 + x_2 \leq 1, \quad -x_1 - x_2 \leq 1.$$

- Each of the four constraints represents one edge of the feasible polytope.
- The constraint functions are chosen so that each one represents a smooth piece of the boundary of Ω.

Smoothness

- In general, the constraint functions are chosen so that each one represents a smooth piece of the boundary of Ω.
- Non-smooth, unconstrained optimization problems can sometimes be reformulated as smooth constrained problems.
- Consider the unconstrained scalar problem of minimizing a non-smooth function $f(x)$ defined by,

$$f(x) = \max(x^2, x)$$

- It has kinks at $x = 0$ and $x = 1$.
- The solution at $x^* = 0$.
- A smooth, constrained formulation of this problem can be obtained by adding an artificial variable t and writing,

Smoothness

$$\min\ t, \quad \text{s.t}, \quad t \geq x, \quad t \geq x^2.$$

- In the examples above we expressed inequality constraints in a slightly different way from the form $c_i(x) \geq 0$.
- However, any collection of inequality constraints with $\geq$ or $\leq$ and nonzero right-hand-sides can be expressed in the form $c_i(x) \geq 0$ by simple rearrangement of the inequality.
- $$t - x \geq 0, \quad t - x^2 \geq 0.$$

EXAMPLES

- To introduce the basic principles behind the characterization of solutions of constrained optimization problems, we work through three simple examples.

Definition

At a feasible point x, the inequality constraint $i \in \mathscr{I}$ is said to be active if $c_i(x) = 0$ and inactive if the strict inequality $c_i > 0$ is satisfied.

Definition

The active set $\mathscr{A}(x)$ at any feasible x consists of the equality constraint indices from $\mathscr{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

$$\mathscr{A}(x) = \mathscr{E} \cup \{i \in \mathscr{I} | c_i(x) = 0\}.$$

Example-1

The first example is a two-variable problem with a single equality constraint:

$$\min\ x_1 + x_2 \qquad x_1^2 + x_2^2 - 2 = 0 \tag{7}$$

- $f(x) = x_1 + x_2$, $\mathscr{I} = \phi$, $\mathscr{E} = \{1\}$
- $c_1(x) = x_1^2 + x_2^2 - 2$
- The feasible set for this problem is the circle of radius $\sqrt{2}$ centered at the origin.
- Just the boundary of this circle, not its interior.
- The solution x^* is $(-1, -1)^T$.

Example-1

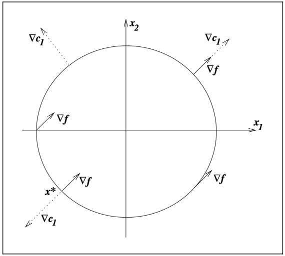

Figure: showing constraint and function gradients at various feasible points.

- From any other point on the circle, it is easy to find a way to move that stays feasible (that is, remains on the circle) while decreasing f.
- From the point $x = (\sqrt{2}, 0)^T$, any move in the clockwise direction around the circle has the desired effect.

A SINGLE EQUALITY CONSTRAINT

- From the figure we see that at the solution x^*, the normal to the constraint $\nabla c_1(x^*)$ is parallel to $\nabla f(x^*)$.
- There is a scalar λ_1^* (in this case $\lambda_1^* = -1/2$) such that

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*). \tag{8}$$

- To retain feasibility with respect to the function $c_1(x) = 0$, it is require for any small (but nonzero) step s to satisfy that $c_1(x + s) = 0$; i.e:

$$0 = c_1(x + s) \approx c_1(x) + \nabla c_1(x)^T s = \nabla c_1(x)^T s.$$

A SINGLE EQUALITY CONSTRAINT

- The step s retains feasibility with respect to c_1, to first order, when it satisfies
$$\nabla c_1(x)^T s = 0. \tag{9}$$
- If we want s to produce a decrease in f;
$$0 > f(x+s) - f(x) \approx \nabla f(x)^T s$$
- or to first order
$$\nabla f(x)^T s < 0 \tag{10}$$

A SINGLE EQUALITY CONSTRAINT

- Existence of a small step s that satisfies both (9) and (10) strongly suggests existence of a direction d where we can get some improvement in the process of minimisation.
- The size of d could be not small; we could have $d \approx s/||s||$ to ensure that the norm of d is close to 1 with the same properties, namely

$$\nabla c_1(x)^T d = 0 \quad \nabla f(x)^T d < 0. \tag{11}$$

- If there is no direction d with the properties (11), then is it likely that we cannot find a small step s with the properties (9) and (10).
- In this case, x^* would appear to be a local minimiser.
- The only way that a d satisfying (11) doesn't exist is if $\nabla f(x)$ and $\nabla c_1(x)$ are parallel.

A SINGLE EQUALITY CONSTRAINT

- Or precisely if the condition

$$\nabla f(x) = \lambda_1 \nabla c_1(x)$$

  holds at x for some scalar λ_1.
- If $\nabla f(x)$ and $\nabla c_1(x)$ are not parallel then we can set:

$$\bar{d} = -\left(\nabla f(x) - \frac{\nabla c_1(x)\nabla f(x)\nabla c_1(x)^T}{||\nabla c_1(x)||^2}\right) \tag{12}$$

  and

$$d = \frac{\bar{d}}{||\bar{d}||} \tag{13}$$

- It can be verified that (13) satisfies (11).

A SINGLE EQUALITY CONSTRAINT

- To write the condition (11) more succinctly we introduce the notion of the *Lagrangian function*.

$$\mathscr{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x). \tag{14}$$

- The gradient w.r.t x of the *Lagrangian* is given by

$$\nabla_x \mathscr{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x) \tag{15}$$

- With the above introduced notions the condition (11) can now be stated as:
  At the solution x^*, there is a scalar λ_1^* such that

$$\nabla_x \mathscr{L}(x^*, \lambda_1^*) = 0. \tag{16}$$

A SINGLE EQUALITY CONSTRAINT

- This observation suggests that we can search for solutions of the equality-constrained problem (7) by seeking stationary points of the Lagrangian function.
- The scalar quantity λ_1 is called a Lagrange multiplier for the constraint $c_1(x) = 0$.
- Though the condition

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$$

  appears to be necessary for an optimal solution of the problem, it is clearly not sufficient.
- The condition is satisfied at the point $x = (1, 1)$ with $\lambda_1 = \frac{1}{2}$.
- But, $(1, 1)$ is obviously not a solution.
- In fact, it maximizes the function f on the circle.

A SINGLE EQUALITY CONSTRAINT

- What may seem a way out from the observation we made in regards to the previous problem is to obtain a sufficient condition for equality-constrained problems is: simply by placing some restriction on the sign of λ_1.
- Consider the constraint

$$x_1^2 + x_2^2 - 2 = 0$$

by its negative i.e.

$$2 - x_1^2 - x_2^2 = 0$$

in the example under consideration.
- The solution of the problem is not affected, but the value of λ_1^* that satisfies the condition (16) changes from $\lambda_1^* = -\frac{1}{2}$ to $\lambda_1^* = \frac{1}{2}$.

A SINGLE INEQUALITY CONSTRAINT

- Here we consider a small modification of Example-1.
- Here the equality constraint is replaced by an inequality.

EXAMPLE-2

Consider

$$\min\ x_1 + x_2 \qquad 2 - x_1^2 - x_2^2 \geq 0 \tag{17}$$

- $f(x) = x_1 + x_2$, $\mathscr{I} = \{1\}$, $\mathscr{E} = \phi$
- $c_1(x) = 2 - x_1^2 - x_2^2$
- The feasible region for this problem is the circle of radius $\sqrt{2}$ centered at the origin.
- Just not the boundary of this circle, but its interior as well.

A SINGLE INEQUALITY CONSTRAINT

- The solution x^* is still $(-1,-1)^T$.
- And the Lagrange multiplier condition holds at $(-1,-1)$ for the value of $\lambda_1^* = \frac{1}{2}$.
- However, this inequality-constrained problem differs from the equality-constrained problem.
- The sign of the Lagrange multiplier plays a significant role, as we now argue.
- Let us conjecture that a given feasible point x is not optimal if we can find a small step s that both retains feasibility and decreases the objective function f to first order.
- The main difference between problems with inequality constraint and equality constraint comes in the handling of the feasibility condition.

A SINGLE INEQUALITY CONSTRAINT

- The step s improves the objective function, to first order, if
$$\nabla f(x)^T s < 0.$$
- s retains feasibility if
$$0 \leq c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s.$$
- That is to first order, feasibility is retained if
$$c_1(x) + \nabla c_1(x)^T s \geq 0. \tag{18}$$
- In determining whether a step s exists that satisfies both the conditions, we consider the following two cases,

A SINGLE INEQUALITY CONSTRAINT

In determining whether a step s exists that satisfies both the conditions, we consider the following two cases,

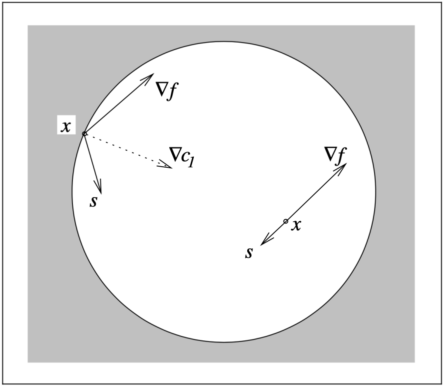

Figure: Improvement directions s from two feasible points x for the problem at which the constraint is active and inactive, respectively

A SINGLE INEQUALITY CONSTRAINT

CASE-1

- Consider first the case in which x lies strictly inside the circle.
- the strict inequality $c_1(x) > 0$ holds.
- In this case, any step vector s satisfies the condition (18), provided only that its length is sufficiently small.
- In fact, whenever $\nabla f(x) \neq 0$, we can obtain a step s that satisfies both the conditions (10) and (18).
- Precisely

  $$s = -\alpha \nabla f(x),$$

  for any positive scalar α sufficiently small.
- This definition does not give a step s with the required properties when

  $$\nabla f(x) = 0$$

A SINGLE INEQUALITY CONSTRAINT

CASE-2

- Consider now the case in which x lies on the boundary of the circle.
- So that $c_1(x) = 0$.
- The conditions (10) and (18) therefore become:

$$\nabla f(x)^T s < 0, \qquad \nabla c_1(x)^T s \geq 0. \tag{19}$$

- The first of these conditions defines an open half-space.
- While the second defines a closed half-space.

A SINGLE INEQUALITY CONSTRAINT

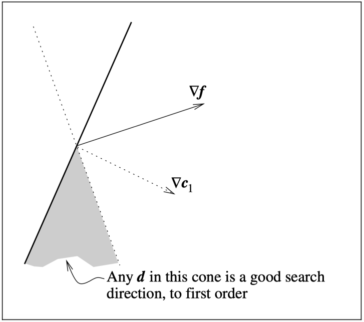

Figure: A direction d that satisfies both conditions (10) and (18) lies in the intersection of a closed half-plane and an open half-plane.

- It is clear from this figure that the intersection of these two regions is empty only when $\nabla f(x)$ and $\nabla c_1(x)$ point in the same direction.

A SINGLE INEQUALITY CONSTRAINT

- That is, when

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \qquad \text{for some } \lambda_1 \geq 0. \tag{20}$$

- The sign of the multiplier is significant here.
- If the Lagrange multiplier condition were satisfied with a negative value of λ_1, then $\nabla f(x)$ and $\nabla c_1(x)$ would point in opposite directions.
- We see from the figure that the set of directions that satisfy both conditions (10) and (18) would make up an entire open half-plane.

A SINGLE INEQUALITY CONSTRAINT

- The optimality conditions for both cases I and II can again be summarized neatly with reference to the Lagrangian function $\mathscr{L}$.
- When no first-order feasible descent direction exists at some point x^*, we have

$$\nabla_x \mathscr{L}(x^*, \lambda_1^*) = 0, \qquad \text{for some } \lambda_1^* \geq 0, \tag{21}$$

- where we also require that

$$\lambda_1^* c_1(x^*) = 0 \tag{22}$$

- Condition (22) is known as a complementarity condition

A SINGLE INEQUALITY CONSTRAINT

- It implies that the Lagrange multiplier λ_1 can be strictly positive only when the corresponding constraint c_1 is active.
- In case I, we have that $c_1(x^*) > 0$.
- So (22) requires that
$$\lambda_1^* = 0$$
- And (21) reduces to
$$\nabla f(x^*) = 0$$
as was required by Case-I.
- In case-II, (22) allows λ_1^* to take on a non-negative value, so (21) becomes equivalent to (20).