Constrained optimization

桂昊坤
HITSZ

2023 年 1 月 13 日

Intro

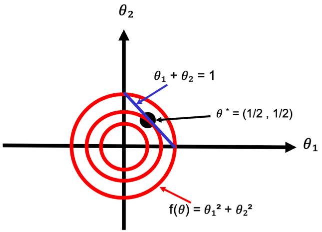
constrained optimization problem

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \mathcal{C}} \mathcal{L}(\boldsymbol{\theta})$$

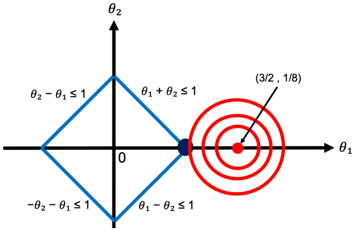
$$\text{where } \mathcal{C} = \{\boldsymbol{\theta} \in \mathbb{R}^D : h_i(\boldsymbol{\theta}) = 0, i \in \mathcal{E}, g_j(\boldsymbol{\theta}) \leq 0, j \in \mathcal{I}\}$$

We call $\mathcal{E}$ the set of **equality constraints**, $\mathcal{I}$ the set of **inequality constraints**

Example:



θ_2
$\theta_1 + \theta_2 = 1$
$\theta^* = (1/2 , 1/2)$
θ_1
$f(\theta) = \theta_1^2 + \theta_2^2$



θ_2
(3/2 , 1/8)
$\theta_2 - \theta_1 \leq 1$
$\theta_1 + \theta_2 \leq 1$
θ_1
0
$-\theta_2 - \theta_1 \leq 1$
$\theta_1 - \theta_2 \leq 1$

Introduction

lemma

1. For any point on the constraint surface, $\nabla_h(\boldsymbol{\theta})$ will be orthogonal to the constraint surface
2. The point $\boldsymbol{\theta}^*$ on the constraint surface such that $\mathcal{L}(\boldsymbol{\theta})$ is minimized satisfies $\nabla\mathcal{L}(\boldsymbol{\theta}^*) = \lambda^*\nabla h(\boldsymbol{\theta}^*)$

Thus, we can construct objective function, which is known as **Lagrangian**.

Lagrangian

$$L(\boldsymbol{\theta}, \lambda) \triangleq \mathcal{L}(\boldsymbol{\theta}) + \lambda h(\boldsymbol{\theta})$$

The constant λ^* is called a **Lagrange multiplier**.

At a stationary point of the Lagrangian, we have

$$\nabla_{\boldsymbol{\theta},\lambda} L(\boldsymbol{\theta}, \lambda) = 0 \Leftrightarrow \lambda \nabla_{\boldsymbol{\theta}} h(\boldsymbol{\theta}) = \nabla \mathcal{L}(\boldsymbol{\theta}), h(\boldsymbol{\theta}) = 0$$

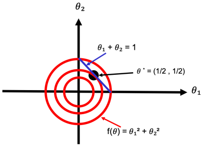
This is called a **critical point**

If we have m > 1 constraints, we can form a new constraint function by addition.

$$L(\boldsymbol{\theta}, \lambda) = \mathcal{L}(\boldsymbol{\theta}) + \sum_{j=1}^{m} \lambda_j h_j(\boldsymbol{\theta})$$

Example

Consider minimizing $\mathcal{L}(\boldsymbol{\theta}) = \theta_1^2 + \theta_2^2$ subject to the constraint that $\theta_1 + \theta_2 = 1$.



The Lagrangian is
$L(\theta_1, \theta_2, \lambda) = \theta_1^2 + \theta_2^2 + \lambda(\theta_1 + \theta_2 - 1)$
By calculating $\nabla_{\theta_1,\theta_2,\lambda} L(\theta_1, \theta_2, \lambda) = \mathbf{0}$, we can get $\boldsymbol{\theta}^* = (0.5, 0.5)$

We generalize the concept of Lagrange multipliers to additionally handle inequality constraints.

First consider the case where we have a single inequality constraint $g(\boldsymbol{\theta}) \leq 0$, thus we can convert it into an unconstrained problem by adding the penalty as an infinite step function.

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \infty \mathbb{I}(g(\boldsymbol{\theta}) > 0)$$

Thus, we can create the following Lagrangian to solve the problem:

$$L(\boldsymbol{\theta}, \mu) = \mathcal{L}(\boldsymbol{\theta}) + \mu g(\boldsymbol{\theta})$$

The step function can be recovered using:

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \max_{\mu \geq 0} L(\boldsymbol{\theta}, \mu) = \begin{cases} \infty & \text{if } g(\boldsymbol{\theta}) > 0 \\ \mathcal{L}(\boldsymbol{\theta}) & \text{otherwise} \end{cases}$$

Thus our optimization problem becomes:

$$\min_{\boldsymbol{\theta}} \max_{\mu \geq 0} L(\boldsymbol{\theta}, \mu)$$

Generalized Lagrangian

Now consider the general case where we have multiple inequality constraints, $g(\boldsymbol{\theta}) \leq 0$, and multiple equality constraints, $h(\boldsymbol{\theta}) = 0$.

generalized Lagrangian

$$L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \mathcal{L}(\boldsymbol{\theta}) + \sum_i \mu_i g_i(\boldsymbol{\theta}) + \sum_j \lambda_j h_j(\boldsymbol{\theta})$$

Our optimization problem becomes:

$$\min_{\boldsymbol{\theta}} \max_{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}} L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

Lagrange dual problem

Lagrange dual problem

The Lagrange dual function is:

$$g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

The corresponding dual problem is:

$$\max_{\boldsymbol{\mu}, \boldsymbol{\lambda}} g(\boldsymbol{\mu}, \boldsymbol{\lambda})$$

$$\text{subject to } \boldsymbol{\mu} \geq \mathbf{0}$$

The Lagrange dual has the following properties:

1. The Lagrange dual function is concave function
   Hint.
   When we fix $\boldsymbol{\theta}$, Lagrange dual function is a Affine function.
2. $\max_{\boldsymbol{\mu},\boldsymbol{\lambda}} g(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq p^*$ where p^* is the optimal value of the primal problem.

Weak duality

Weak duality

$$\max_{\boldsymbol{\mu},\boldsymbol{\lambda}} \min_{\boldsymbol{\theta}} L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) \leq \min_{\boldsymbol{\theta}} \max_{\boldsymbol{\mu},\boldsymbol{\lambda}} L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda})$$

Proof.

$$f(\boldsymbol{\mu},\boldsymbol{\lambda}) = \min_{\boldsymbol{\theta}} L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) \leq L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) \leq \max_{\boldsymbol{\mu},\boldsymbol{\lambda}} L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) = g(\boldsymbol{\theta})$$

$$\max_{\boldsymbol{\mu},\boldsymbol{\lambda}} f(\boldsymbol{\mu},\boldsymbol{\lambda}) \leq \min_{\boldsymbol{\theta}} g(\boldsymbol{\theta}) \qquad \square$$

We can also get the Geometric View of Weak Duality:
Let us assume for simplicity that m = 1, and consider the problem:

$$\min_x f_0(x) : f_1(x) \leq 0$$

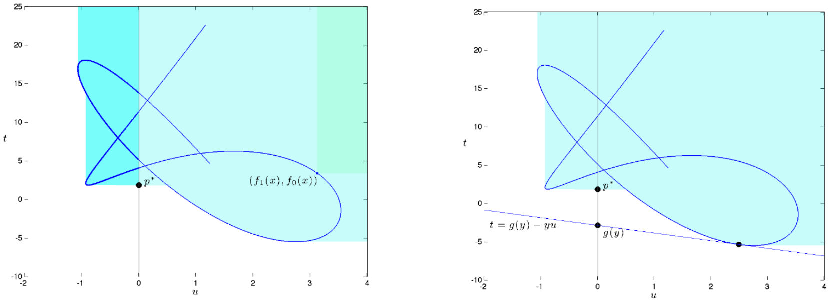
We can express the primal problem with two new scalar variables u,t, as follows: The line in blue is the set of points of the form $(u, t) = (f_1(x), f_0(x))$.
The dual function is:

$$g(y) = \min_x f_0(x) + yf_1(x)$$

Then we have:

$$g(y) = \min_{u,t} t + yu : (u, t) \in A$$

Weak duality



where $\mathbf{A} := \{(u, t) \in \mathbf{R}^2 : \exists x, u \geq f_1(x), t \geq f_0(x)\}$

Strong duality

The theory of weak duality seen here states that $p^* \geq d^*$. This is true always, even if the original problem is not convex.

Strong duality

Strong duality holds if $p^* = d^*$.

So when could we get Strong duality?

Slater condition

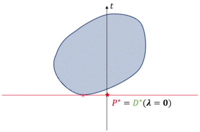
Slater condition

- The primal problem is convex;
- It is strictly feasible, that is, there exists $x_0 \in \mathbf{R}^n$ such that

$$Ax_0 = b, f_i(x_0) < 0, i = 1, \ldots, m$$

Then, strong duality holds: $p^* = d^*$, and the dual problem is attained.

Slater condition



Primal problem is convex, that is, f_0 and f_1 are convex functions, the above set is **convex**.

Primal feasibility means that the set $\mathbf{A}$ cuts "inside" the right-half of the (u,t)-plane.

KKT conditions

- Stationarity: $\nabla \mathcal{L}(\boldsymbol{\theta}^*) + \sum_i \mu_i g_i(\boldsymbol{\theta}) + \sum_j \lambda_j h_j(\boldsymbol{\theta}) = \mathbf{0}$
- Primal feasibility: $g(\boldsymbol{\theta}) \leq 0, h(\boldsymbol{\theta}) = \mathbf{0}$
- Dual feasibility: $\boldsymbol{\mu} \geq \mathbf{0}$
- Complementary slackness: $\boldsymbol{\mu} \odot \boldsymbol{g} = \mathbf{0}$

Theorem

For a problem with strong duality, x^* and u^*, v^* satisfy the KKT conditions if and only if x^* and u^*, v^* are primal and dual solutions.

Theorem

Proof.

Necessity:

$$
\begin{aligned}
f(x^*) &= g(u^*, v^*) \\
&= \min_x f(x) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{j=1}^{r} v_j^* l_j(x) \\
&\le f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* l_j(x^*) \\
&\le f(x^*)
\end{aligned}
$$

Theorem

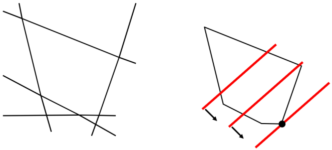
Sufficiency:

$$
\begin{aligned}
g(u^*, v^*) &= f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* l_j(x^*) \\
&= f(x^*)
\end{aligned}
$$

Linear programming

Consider optimizing a linear function subject to linear constraints.

$$\min_{\boldsymbol{\theta}} \mathbf{c}^T\boldsymbol{\theta} \quad s.t. \quad \mathbf{A}\boldsymbol{\theta} \leq \mathbf{b}, \boldsymbol{\theta} \geq \mathbf{0}$$



(a) (b)

The simplex algorithm

The **simplex algorithm** solves LPs by moving from vertex to vertex, each time seeking the edge which most improves the objective.

In the worst-case scenario, the simplex algorithm can take time exponential in D, although in practice it is usually very efficient. There are also various polynomial-time algorithms, such as the interior point method, although these are often slower in practice.

Quadratic programming

Consider minimizing a quadratic objective subject to linear equality and inequality constraints.

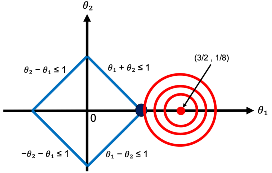
$$\min_{\boldsymbol{\theta}} \frac{1}{2}\boldsymbol{\theta}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{c}^T\boldsymbol{\theta} \quad s.t. \quad \mathbf{A}\boldsymbol{\theta} \leq \mathbf{b}, \mathbf{C}\boldsymbol{\theta} = \mathbf{d}$$

If $\mathbf{H}$ is positive semidefinite, then this is a convex optimization problem.

Suppose we want to minimize:

$$\mathcal{L}(\boldsymbol{\theta}) = (\theta_1 - \frac{3}{2})^2 + (\theta_2 - \frac{1}{8})^2 = \frac{1}{2}\boldsymbol{\theta}^T \mathbf{H}\boldsymbol{\theta} + \mathbf{c}^T\boldsymbol{\theta} + \text{const}$$

where $\mathbf{H} = 2\mathbf{I}$ and $\mathbf{c} = -(3, \frac{1}{4})$, subject to $|\theta_1| + |\theta_2| \leq 1$



$g_3(\boldsymbol{\theta}^*) > 0$ and $g_4(\boldsymbol{\theta}^*) > 0$, and hence, by complementarity, $\mu_3^* = \mu_4^* = 0$. We can therefore remove these inactive constraints.

Hence the solution is $\boldsymbol{\theta}^* = (1, 0)^T$,$\boldsymbol{\mu}^* = (0.625, 0.375, 0, 0)^T$

Mixed integer linear programming

Integer linear programming or **ILP** corresponds to minimizing a linear objective, subject to linear constraints, where the optimization variables are discrete integers instead of reals.

$$\min_{\boldsymbol{\theta}} \mathbf{c}^T\boldsymbol{\theta} \quad s.t. \quad \mathbf{A}\boldsymbol{\theta} \leq \mathbf{b}, \boldsymbol{\theta} \geq \mathbf{0}, \boldsymbol{\theta} \in \mathbb{Z}^D$$

If some of the optimization variables are real-valued, it is called a **mixed ILP**, often called a **MIP** for short.

Reference

- Slater Condition for Strong Duality
- KKT conditions