AI505

Optimization

Constrained Optimization

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Constrained Optimization

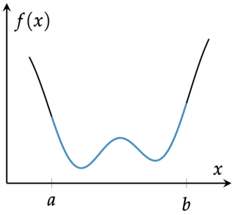
- Minimizing an objective subject to design point restrictions called **constraints**
- A variety of techniques transform constrained optimization problems into unconstrained problems
- New optimization problem statement

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned}$$

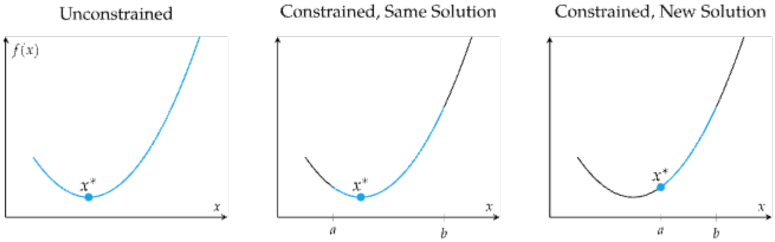
- The set $\mathcal{X} \subset \mathbb{R}$ is called the **feasible set**

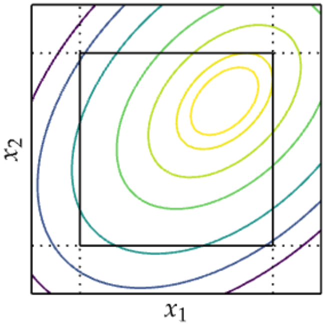
Constrained Optimization

Constraints that bound feasible set can change the optimizer



$$\underset{x}{\text{minimize}}\ f(x)$$
$$\text{subject to } x \in [a, b]$$





Constraint Types

- Generally, constraints are formulated using two types:
    1. **Equality constraints**: $h(\boldsymbol{x}) = 0$
    2. **Inequality constraints**: $g(\boldsymbol{x}) \leq 0$
- Any optimization problem can be written as

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq 0 \text{ for all } i \text{ in } \{1, \ldots, m\} \\ & h_j(\boldsymbol{x}) = 0 \text{ for all } j \text{ in } \{1, \ldots, \ell\} \end{aligned}$$

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{aligned}$$

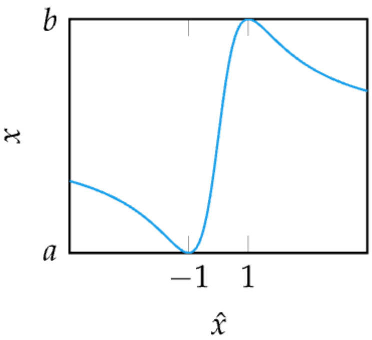
f and the functions h and g are all smooth, real-valued functions on a subset of Re^n

Transformations to Remove Constraints

- If necessary, some problems can be reformulated to incorporate constraints into the objective function

- If x is constrained between a and b

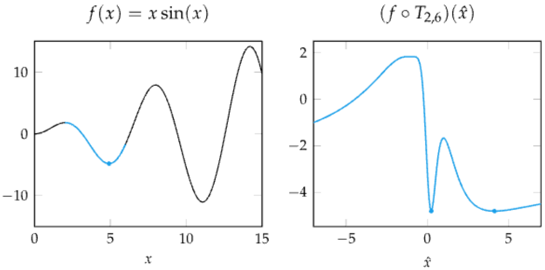
$$x = t_{a,b}(\hat{x}) = \frac{b+a}{2} + \frac{b-a}{2}\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)$$



Transformations to Remove Constraints

Example

$$\underset{x}{\text{minimize}} \; x \sin(x)$$
$$\text{subject to } 2 \leq x \leq 6$$



$$\underset{\hat{x}}{\text{minimize}} \; t_{2,6}(\hat{x}) \sin(t_{2,6}(\hat{x}))$$

$$\underset{\hat{x}}{\text{minimize}} \; \left(4 + 2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right) \sin\left(4 + 2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right)$$

Transformations to Remove Constraints

Example

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h(\boldsymbol{x}) = x_1^2 + x_2^2 + \ldots + x_n^2 - 1 = 0 \end{aligned}$$

- Solve for one of the variables to eliminate constraint:

$$x_n = \pm\sqrt{1 - x_1^2 - x_2^2 - \ldots - x_{n-1}^2}$$

- Transformed, unconstrained optimization problem:

$$\underset{x}{\text{minimize}} \quad \left(\left[x_1, x_2, \ldots, x_{n-1}, \pm\sqrt{1 - x_1^2 - x_2^2 - \ldots - x_{n-1}^2}\right]\right)$$

Lagrangian Relaxation

- With only equality constraints, critical points (local minima, global minima, or saddle points optimal) where gradient of f and the gradient of h are aligned

- The method of Lagrangian relaxation is used to optimize a function subject to (equality) constraints

- Lagrangian multipliers refer to the variables introduced by the method denoted by λ

$$\underset{\boldsymbol{x}}{\text{minimize}} \;\; f(\boldsymbol{x})$$
$$\text{subject to} \;\; h(\boldsymbol{x}) = 0$$

1. Form **Lagrangian relaxation**

$$\mathcal{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda h(\boldsymbol{x})$$

2. Set $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) = 0$ and $\nabla_{\lambda} \mathcal{L}(\boldsymbol{x}, \lambda) = 0$ to get

$$\nabla f(\boldsymbol{x}) = \lambda \nabla h(\boldsymbol{x}) \qquad h(\boldsymbol{x}) = 0$$

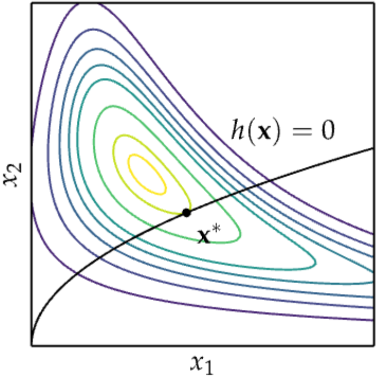
3. solve for $\boldsymbol{x}$ and λ

Example

$$
\begin{aligned}
&\text{minimize} \quad -\exp\left(-\left(x_1 x_2 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{3}{2}\right)^2\right)\\
&\text{subject to } x_1 - x_2^2 = 0
\end{aligned}
$$

Lagrangian Relaxation

Intuitively, the method of Lagrange multipliers finds the point x^* where the constraint function is orthogonal to the gradient



Lagrangian Relaxation with Inequality Constraints

$$\underset{\boldsymbol{x}}{\text{minimize}} \; f(\boldsymbol{x})$$
$$\text{subject to } g(\boldsymbol{x}) \leq 0$$

- If solution lies at the constraint boundary, the constraint is called **active**, and the Lagrangian condition holds for a non-negative constant μ:

$$\nabla f(\boldsymbol{x}) + \mu \nabla g(\boldsymbol{x}) = 0$$

- If the solution lies within the boundary, the constraint is called **inactive**, and the optimal solution simply lies where

$$\nabla f(\boldsymbol{x}) = 0$$

that is, the Lagrangian condition holds with $\mu = 0$

Lagrangian Relaxation with Inequality Constraints

$$\underset{\boldsymbol{x}}{\text{minimize}} \;\; f(\boldsymbol{x})$$
$$\text{subject to} \;\; g(\boldsymbol{x}) \leq 0$$

- We create the Lagrangian relaxation such that it goes to ∞ outside the feasibility set ($g(\boldsymbol{x}) \not\leq 0$)):

$$\mathcal{L}_{\infty}(\boldsymbol{x}) = f(\boldsymbol{x}) + \infty(g(\boldsymbol{x}) > 0)$$

impractical: discontinuous and nondifferentiable.

- Instead, for $\mu > 0$:

$$\mathcal{L}(\boldsymbol{x}, \mu \geq 0) = f(\boldsymbol{x}) + \mu g(\boldsymbol{x})$$

$$\mathcal{L}_{\infty}(\boldsymbol{x}) = \underset{\mu \geq 0}{\text{maximize}} \; \mathcal{L}(\boldsymbol{x}, \mu)$$

for $\boldsymbol{x}$ infeasible, $\mathcal{L}_{\infty}(\boldsymbol{x}) = \infty$; for $\boldsymbol{x}$ feasible, $\mathcal{L}_{\infty}(\boldsymbol{x}) = f(\boldsymbol{x})$

- The new optimization problem becomes

$$\underset{\boldsymbol{x}}{\text{minimize}} \, \underset{\mu \geq 0}{\text{maximize}} \; \mathcal{L}(\boldsymbol{x}, \mu)$$

This is called the **primal problem**

Necessary Conditions – KKT Conditions

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{aligned}$$

Any critical point $\boldsymbol{x}^*$ must satisfy the **Karush-Kuhn-Tucker conditions**

1. primal feasibility: $\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$ and $\boldsymbol{h}(\boldsymbol{x}^*) = 0$
2. dual feasibility: penaliztion is towards feasibility $\boldsymbol{\mu} \geq 0$
3. complementary slackness: either μ_i or $g_i(\boldsymbol{x}^*)$ is zero.

   $$\mu_i g_i(\boldsymbol{x}^*) = 0, \text{ for } i = 1, \ldots, m.$$

4. stationarity: objective function tanget to each active constraint

   $$\nabla f(\boldsymbol{x}^*) + \sum_i \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_j \lambda_j \nabla h_j(\boldsymbol{x}^*) = 0$$

Necessary Conditions – KKT Conditions

Particular cases

- f concave, $\boldsymbol{g}$ convex: then KKT are also sufficient
- Patological cases

In vector form:

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \boldsymbol{\mu} \nabla \cdot \boldsymbol{g}(\boldsymbol{x}^*) + \boldsymbol{\lambda} \cdot \boldsymbol{h}(\boldsymbol{x}^*) = 0 \\ \boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x}^*) = 0 \\ \boldsymbol{g}(\boldsymbol{x}^*) \leq 0, \ \boldsymbol{h}(\boldsymbol{x}^*) = 0 \\ \boldsymbol{\mu} \geq 0 \end{cases}$$

Duality

- Generalized Lagrangian Relaxation:

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

- the primal form is

$$\underset{\boldsymbol{x}}{\text{minimize}} \underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}} \; \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}.\boldsymbol{\lambda})$$

- Reversing the order of operations leads to the **dual form**

$$\underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}} \underset{\boldsymbol{x}}{\text{minimize}} \; \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

- In some cases, the dual problem is easier to solve computationally than the original problem. In other cases, the dual can be used to obtain easily a lower bound on the optimal value of the objective for the primal problem. The dual has also been used to design algorithms for solving the primal problem.

Duality

Theorem (Max-min inequality)

For any function $f : Z \times W \to \mathbb{R}$,

$$\sup_{z \in Z} \inf_{w \in W} f(z, w) \leq \inf_{w \in W} \sup_{z \in Z} f(z, w) \ .$$

Proof: see wikipedia $\square$

- When f, W, and Z are convex the inequality becomes equality and we have a strong max–min property (or a saddle-point property).
- For us:

$$\underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}} \, \underset{\boldsymbol{x}}{\text{minimize}} \, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq \underset{\boldsymbol{x}}{\text{minimize}} \, \underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}} \, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}.\boldsymbol{\lambda})$$

- Therefore, the solution to the dual problem d^* is a lower bound to the primal solution p^*
- The inner part of the dual problem can be used to define the **dual function** or **dual objective**

$$\mathcal{D}(\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}) = \underset{\boldsymbol{x}}{\text{minimize}} \, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

Duality

- The dual function is concave. Gradient ascent on a concave function always converges to the global maximum.
- **Dual Problem**: $\max \mathcal{D}(\lambda)$ subject to $\lambda \geq 0$
- Optimizing the dual problem is easy whenever minimizing the Lagrangian with respect to x is easy.
- For any $\boldsymbol{\mu} \geq 0$ and any $\boldsymbol{\lambda}$, we have

$$\mathcal{D}(\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}) \leq p^*$$

- The difference between dual and primal solutions d^* and p^* is called the **duality gap**
- Showing zero-duality gap is a "certificate" of optimality

Penalty methods

- Penalty methods are a way of reformulating a constrained optimization problem as an unconstrained problem by penalizing the objective function value when constraints are violated

Example

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{aligned}$$

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) + \rho \cdot p_{count}(\boldsymbol{x}) \\ \text{s.t.} \quad & p_{count}(\boldsymbol{x}) = \sum_i (g_i(\boldsymbol{x}) > 0) + \sum_j (h_j(\boldsymbol{x}) \neq 0) \end{aligned}$$

Penalty Methods

```
Procedure penalty_method;
Input: f, p, x, k_max; ρ = 1, γ = 2
Output: x  solution
for k in 1, ..., k_max do
    x ← minimize_x{f(x) + ρ · p(x)};
    ρ ← ρ · γ;
    if p(x) = 0 then
        return x;
return x;
```

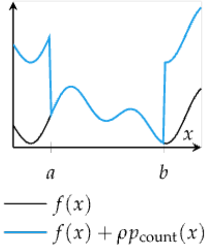
Procedure penalty_method;
Input: $f, p, \boldsymbol{x}, k_{max}; \rho = 1, \gamma = 2$
Output: $\boldsymbol{x}$ solution
for k in $1, ..., k_{max}$ **do**
- $\boldsymbol{x} \leftarrow \text{minimize}_{\boldsymbol{x}}\{f(\boldsymbol{x}) + \rho \cdot p(\boldsymbol{x})\}$;
- $\rho \leftarrow \rho \cdot \gamma$;
- **if** $p(\boldsymbol{x}) = 0$ **then**
  - return $\boldsymbol{x}$;

return $\boldsymbol{x}$;

Penalty methods

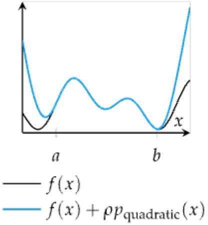
- Count penalty:

$$p_{count}(\boldsymbol{x}) = \sum_i (g_i(\boldsymbol{x}) > 0) + \sum_j (h_j(\boldsymbol{x}) \neq 0)$$



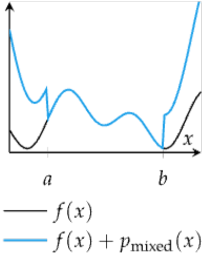
- Quadratic penalty:

$$p_{quadratic}(\boldsymbol{x}) = \sum_i \max(g_i(\boldsymbol{x}), 0)^2 + \sum_j h_j(\boldsymbol{x})^2$$



- Mixed Penalty:

$$p_{mixed}(\boldsymbol{x}) = \rho_1 p_{count}(\boldsymbol{x}) + \rho_2 p_{quadratic}(\boldsymbol{x})$$



Augmented Lagrange Method

- Adaptation of penalty method for equality constraints

$$p_{Lagrangian}(\boldsymbol{x}) \stackrel{def}{=} \frac{1}{2}\rho \sum_i (h_i(\boldsymbol{x}))^2 - \sum_i \lambda_i h_i(\boldsymbol{x})$$

Procedure augmented_lagrange_method;
Input: $f, h, \boldsymbol{x}, k_{max}; \rho = 1, \gamma = 2)$
$\boldsymbol{\lambda} \leftarrow 0$;
for k in $1, \ldots, k_{max}$ **do**
    $p \leftarrow \left(x \mapsto \rho/2 \cdot \sum_i (h_i(\boldsymbol{x})^2) - \boldsymbol{\lambda} \cdot h(\boldsymbol{x})\right)$;
    $\boldsymbol{x} \leftarrow \text{minimize}_{\boldsymbol{x}}\{f(\boldsymbol{x}) + p(\boldsymbol{x})\}$;
    $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} - \rho \cdot \boldsymbol{h}(\boldsymbol{x})$;
    $\rho \leftarrow \rho \cdot \gamma$;
return $\boldsymbol{x}$;

- $\boldsymbol{\lambda}$ converges towards the Lagrangian multiplier

Interior Point Methods

- Also called **barrier methods**, interior point methods ensure that each step is feasible
- This allows premature termination to return a nearly optimal, feasible point
- Barrier functions are implemented similar to penalties but must meet the following conditions:
  1. Continuous
  2. Non-negative
  3. Approach infinity as x approaches boundary

Interior Point Methods

- Inverse Barrier:

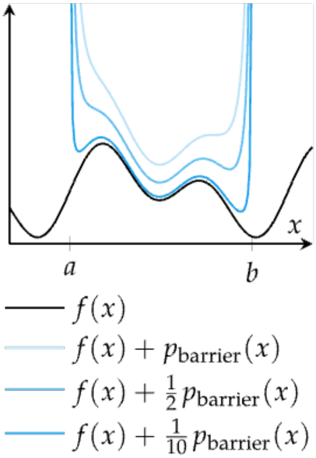
$$p_{barrier}(\boldsymbol{x}) = -\sum_i \frac{1}{g_i(\boldsymbol{x})}$$

- Log Barrier:

$$p_{barrier}(\boldsymbol{x}) = -\sum_i \begin{cases} \log(-g_i(\boldsymbol{x})) & \text{if } g_i(\boldsymbol{x}) \geq -1 \\ 0 & \text{otherwise} \end{cases}$$

New optimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}}\ f(\boldsymbol{x}) + \frac{1}{\rho} p_{barrier}(\boldsymbol{x})$$



Interior Point Methods

Procedure interior_point_method;
Input: $f, p, \boldsymbol{x}; \rho = 1, \gamma = 2, \epsilon = 0.001$
$\Delta \leftarrow \infty$;
while $\Delta > \epsilon$ **do**
    $\boldsymbol{x}' \leftarrow \text{minimize}_{\boldsymbol{x}}\{f(\boldsymbol{x}) + p(\boldsymbol{x})/\rho\}$;
    $\Delta \leftarrow \|\boldsymbol{x}' - \boldsymbol{x}\|$;
    $\boldsymbol{x} \leftarrow \boldsymbol{x}'$;
    $\rho \leftarrow \rho \cdot \gamma$;
return $\boldsymbol{x}$;

- Line searches $f(\boldsymbol{x} + \alpha \boldsymbol{d})$ are constrained to the interval $\alpha = [0, \alpha_u]$, where α_u is the step to the nearest boundary.
  In practice, α_u is chosen such that $\boldsymbol{x} + \alpha \boldsymbol{d}$ is just inside the boundary to avoid the boundary singularity.
- Needs an initial **feasible** solutions. Typically, found by solving:

$$\underset{\boldsymbol{x}}{\text{minimize}}\ p_{quadratic}(\boldsymbol{x})$$

Summary

- Constraints are requirements on the design points that a solution must satisfy
- Some constraints can be transformed or substituted into the problem to result in an unconstrained optimization problem
- Analytical methods using Lagrange multipliers yield the generalized Lagrangian and the necessary conditions for optimality under constraints
- A constrained optimization problem has a dual problem formulation that is easier to solve and whose solution is a lower bound of the solution to the original problem
- Penalty methods penalize infeasible solutions and often provide gradient information to the optimizer to guide infeasible points toward feasibility
- Interior point methods maintain feasibility but use barrier functions to avoid leaving the feasible set