Constrained Optimization I: Introduction

Asst. Prof. Dr.-Ing. Sudchai Boonto

Department of Control System and Instrumentation Engineering
King Mongkut's Unniversity of Technology Thonburi
Thailand

November 4, 2025

Objective

At the end of this chapter you should be able to:

- Describe and implement the constrained optimization problems
- Understand the concept of **Lagrange multipliers**
- Understand the **Karush-Kuhn-Tucker** conditions

Notation and Basic Assumptions

Definition: Constrained Optimization Problem

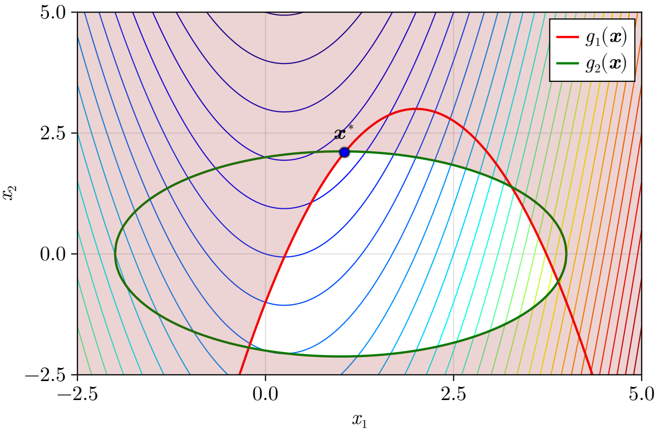
$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1, 2, \ldots, p \\
& g_j(\mathbf{x}) \leq 0 \quad \text{for } j = 1, 2, \ldots, q
\end{aligned}
$$

where $h_i(\mathbf{x})$ is a **equality constraint**, and $g_j(\mathbf{x})$ is the vector of **inequality constraint**.

Consider a two-variable problem

$$
\begin{aligned}
\underset{x_1, x_2}{\text{minimize}} \quad & f(x_1, x_2) = x_1^2 - \frac{1}{2}x_1 - x_2 - 2 \\
\text{subject to} \quad & g_1(x_1, x_2) = x_1^2 - 4x_1 + x_2 + 1 \leq 0 \\
& g_2(x_1, x_2) = \frac{1}{2}x_1^2 + x_2^2 - x_1 - 4 \leq 0
\end{aligned}
$$

Notation and Basic Assumptions



A graphical method can be used to solve simple problems. However, it is difficult or impossible to use such a method for more constrained functions and high-dimensional systems.

Equality Constraints: Direct Substitution

Consider a problem with the next level of complexity: Optimization with equality constraints

$$\begin{aligned} \underset{\mathbf{y}}{\text{minimize}} \quad & f(\mathbf{y}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{y}) = 0 \end{aligned}$$

- To simplify the notation, let be partitioned into a decision vector and a state vector, such that

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}^\top & \mathbf{u}^\top \end{bmatrix}^\top \in \mathbb{R}^p,\ \mathbf{x} \in \mathbb{R}^n,\ \mathbf{u} \in \mathbb{R}^m,\ p = m + n$$

where is implicitly defined by the constraints that relate it to the decision variables. The problem now becomes:

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & f(\mathbf{x}, \mathbf{u}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}, \mathbf{u}) = 0 \end{aligned}$$

Equality Constraints: Direct Substitution

- p must be greater than n otherwise the problem is completely specified by the constraints (or over specified or not depend of f).
- One solution approach to solve the problem is direct substitution, which involves
  - Solving for $\mathbf{x}$ in terms of $\mathbf{u}$ using $\mathbf{h}(\mathbf{x}, \mathbf{u})$
  - Substituting this expression into $f(\mathbf{x}, \mathbf{u})$ and solving for $\mathbf{u}$ using an unconstrained optimization.
  - The method is good if $f(\mathbf{x}, \mathbf{u})$ is linear (assumption is that not both of f and $\mathbf{h}$ are linear.)

Equality Constraints: Direct Substitution

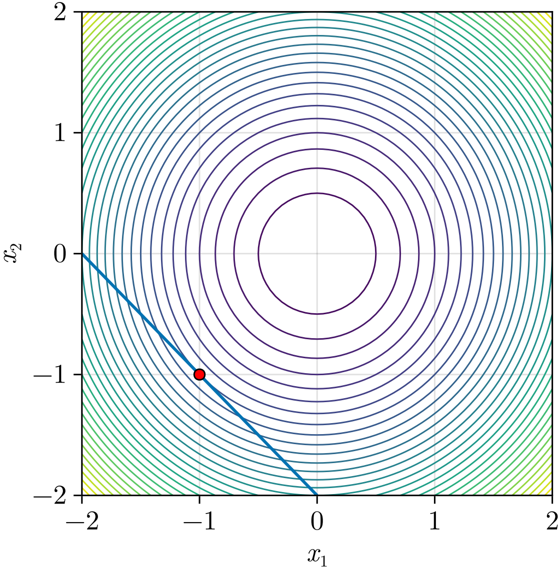
$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) = x_1^2 + x_2^2$$
$$\text{subject to} \quad x_1 + x_2 + 2 = 0$$

- Clearly the unconstrained minimum is at $x_1 = x_2 = 0$
- From the constrain, $x_1 = -2 - x_2$ or $x_2 = -2 - x_1$, we have equivalent problems:

$$\underset{x_1}{\text{minimize}} \quad f_1(x_1) = x_1^2 + (-2 - x_1)^2 \text{ or } \underset{x_2}{\text{minimize}} \quad f_2(x_2) = (-2 - x_2)^2 + x_2^2$$

- the solution ($\partial f_i / \partial x_i = 0$ is $x_1 = x_2 = -1$
- The substitution method works well for linear constraints, but it is hard to generalize for larger systems/ nonlinear constraints.

Equality Constraints: Direct Substitution



Notation and Basic Assumptions

- For unconstrained gradient-based optimization, we only require the gradient of the objective, $\nabla f(\mathbf{x})$. To solve a constrained problem, we also require the gradients of all the constraints. Because the constraints are vectors, their derivatives yield a **Jacobian** matrix. For the equality constraints, we have

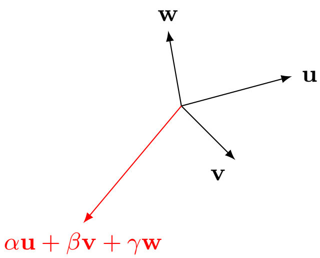
$$\mathbf{J_h} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \underbrace{\begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_p}{\partial x_1} & \cdots & \frac{\partial h_p}{\partial x_n} \end{bmatrix}}_{p \times n} = \begin{bmatrix} \nabla h_1^\top \\ \vdots \\ \nabla h_p^\top \end{bmatrix}$$

- Similarly, the Jacobian of the inequality constraints is an $(q \times n)$ matrix.

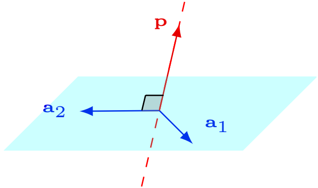
n-dimension space

There are several essential linear algebra concepts for constrained optimization.

- The **span** of a set of vectors is the space formed by all points that can be obtained by a linear combination of those vectors.
- The **null space** of a matrix $\mathbf{A}$ is the set of all n-dimensional vector $\mathbf{p}$ such that $\mathbf{A}\mathbf{p} = 0$.

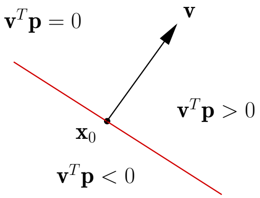


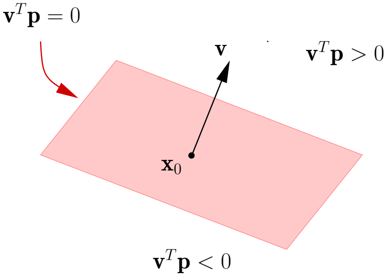
Span in three-dimensional space.



Nullspace of a 2×3 matrix A of rank 2, where a_1 and a_2 are the row vectors of A.

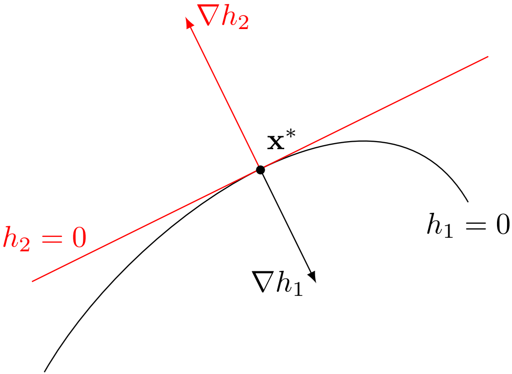
Hyperplanes and Half-space





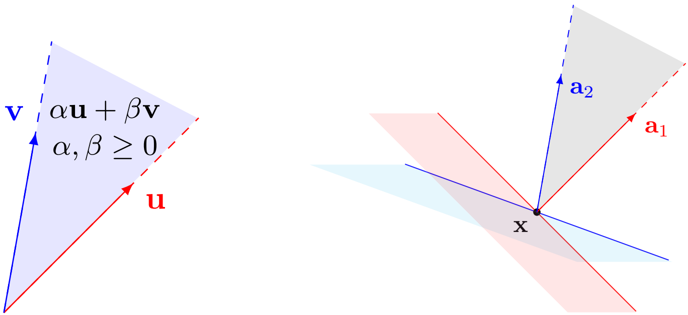
- In n dimensions, a hyperplane of $n-1$ dimensions divides the space into two **half-spaces**: in one of these, $\mathbf{v}^\top\mathbf{p} > 0$, and in the other, $\mathbf{v}^\top\mathbf{p} < 0$.
- Each half-space is closed if it includes the hyperplane ($\mathbf{v}^\top\mathbf{p} = 0$) and open otherwise.

Hyperplanes and Half-space



- The function gradient at the point on the isosurface is locally perpendicular to the isosurface. The gradient vector defines the **tangent hyperplane** and the point.
- The set of points such that $\nabla f^\top p = 0$.

Hyperplanes and Half-space



- The intersection of multiple half-spaces yields a **polyhedral cone.**
- A polyhedral cone is the set of all the points that can be obtained by the linear combination of a given set of vectors using nonnegative coefficients.

Equality Constraints

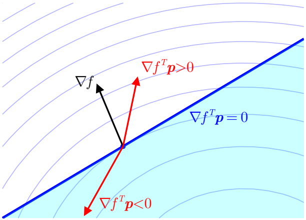
For the unconstrained case, by taken a first-order Taylor series expansion of the objective function with some step $\mathbf{p}$ that is small enough by neglecting the second-order term:

$$f(\mathbf{x}+\mathbf{p}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{p}$$

At the minimum point $\mathbf{x}^*$, we should have

$$f(\mathbf{x}^* + \mathbf{p}) \geq f(\mathbf{x}^*) \qquad \Rightarrow \qquad \nabla f(\mathbf{x}^*)^\top \mathbf{p} \geq 0$$

For unconstraint problem, $\nabla f^\top \mathbf{p} \geq 0$ is satisfied if $\nabla f(\mathbf{x}^*) = 0$



The gradient $f(\mathbf{x})$, which is the direction of steepest function increase, splits the design space into two halves. All $\mathbf{p}$ direction that make the function decrease always make $\nabla f^\top \mathbf{p} < 0$ except when $\nabla f^\top \mathbf{p} = 0$.

Equality Constraints

- For constrained problem, the function increase condition still applies, but $\mathbf{p}$ must also be a **feasible** direction. To find the feasible directions, we use a first-order Taylor series expansion for each equality constraint function as

$$h_j(\mathbf{x}+\mathbf{p}) \approx h_j(\mathbf{x}) + \nabla h_j(\mathbf{x})^\top \mathbf{p}, \qquad j = 1, \ldots, p$$

- $\mathbf{x}$ is a feasible point, then $h_j(\mathbf{x}) = 0$ for all constraints j, then

$$\nabla h_j(\mathbf{x})^\top \mathbf{p} = 0, \qquad \text{for all } j = 1, \ldots, p$$

- The **direction $\mathbf{p}$ is feasible when it is orthogonal to all equality constraint gradients.** Or,

$$\mathbf{J}_h(\mathbf{x})\mathbf{p} = 0$$

- Any feasible direction has to lie in the nullspace of the Jacobian of the constraints, $\mathbf{J}_h$.

Equality Constraints

- For constrained optimality, we need to satisfy both $\nabla f(\mathbf{x}^*)^\top \mathbf{p} \geq 0$ and $\mathbf{J}_h(\mathbf{x})\mathbf{p} = 0$
- For equality constraints, if a direction $\mathbf{p}$ is feasible, then $-\mathbf{p}$ must also be feasible (from Taylor series), Therefore, the only way to satisfy $\nabla f(\mathbf{x}^*)^\top \mathbf{p} \geq 0$ is if $\nabla f(\mathbf{x})^\top \mathbf{p} = 0$.

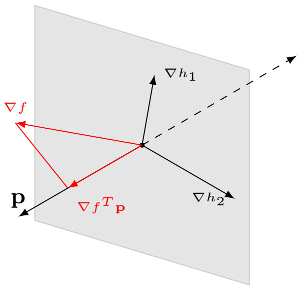
Theorem: 1^{st} order condition

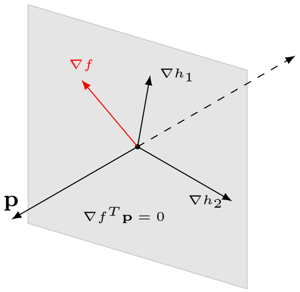
For $\mathbf{x}^*$ to be constrained optimum, we require

$$\nabla f(\mathbf{x}^*)^\top \mathbf{p} = 0 \quad \text{for all } \mathbf{p} \text{ such that} \quad \mathbf{J}_h(\mathbf{x}^*)\mathbf{p} = 0$$

- On other words, **the projection of the objective function gradient onto the feasible space must vanish.**

Equality Constraints





- The objective function gradient must be a linear combination of the gradients of the constraints. (left) we still have decent direction. (right) $\mathbf{x}$ is optimal.
- We can write

$$\nabla f(\mathbf{x}^*) = -\sum_{j=1}^{p} \lambda_j \nabla h_j(\mathbf{x}^*)$$

- λ_j are called the **Lagrange multipliers**. For equality constraints, the sign of Lagrange multipliers is arbitrary.

Equality Constraints

It is more convenient to use the **Lagrangian function:**

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{h}(\mathbf{x})$$

$$\nabla_\mathbf{x} \mathcal{L} = \nabla f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{J}_h(\mathbf{x}) = 0, \qquad \nabla_{\boldsymbol{\lambda}} \mathcal{L} = \mathbf{h}(\mathbf{x}) = 0$$

With the Lagrangian function, we have transformed a constrained problem into an unconstrained problem by introducing new variables, $\boldsymbol{\lambda}$.

Theorem: 1^{st}-order optimality conditions

The optimality conditions for the equality-constrained case are

$$\begin{aligned} \nabla f(\mathbf{x}^*) + (\boldsymbol{\lambda}^*)^\top \mathbf{J}_h(\mathbf{x}^*) &= 0 \\ \mathbf{h}(\mathbf{x}^*) &= 0 \end{aligned}$$

These conditions assume that the gradients of the constraints are linearly independent; that is, $\mathbf{J}_h$ has full row rank.

Equality Constraints

The set of equality constraints

$$h_1(\mathbf{x}) = 0, h_2(\mathbf{x}) = 0, \cdots, h_p(\mathbf{x}) = 0$$

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) & h_2(\mathbf{x}) & \cdots & h_p(\mathbf{x}) \end{bmatrix}^\top, \mathbf{h}(\mathbf{x}) = 0$$

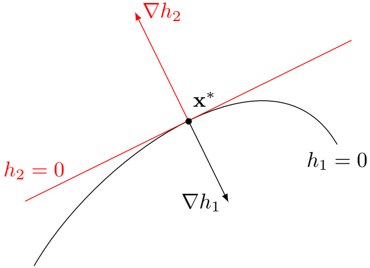
Definition: Regular point

A point $\mathbf{x}$ is called a **regular point** of the constraints $\mathbf{h}(\mathbf{x})$ if $\mathbf{x}$ satisfies $\mathbf{h}(\mathbf{x}) = 0$ and column vectors $\nabla h_1(\mathbf{x})$, $\nabla h_2(\mathbf{x})$, $\cdots$, $\nabla h_p(\mathbf{x})$ are linearly independent.

- The definition states that $\mathbf{x}$ is a regular point of the constraints if it is a solution of $\mathbf{h}(\mathbf{x}) = 0$ and the Jacobian $\mathbf{J}_h = \begin{bmatrix} \nabla h_1(\mathbf{x}) & \nabla h_2(\mathbf{x}) & \cdots & \nabla h_p(\mathbf{x}) \end{bmatrix}^\top$
- It is impossible for $\mathbf{x}$ to be a regular point of the constraints if $p > n$. It is the upper bound for the number of independent equality constraints, i.e., $p \leq n$.

Equality Constraints

- For $p > n$, the constraint qualification condition does not hold in this case because the gradients of the two constraints not linearly independent.



- The optimality conditions using first-oder conditions is a necessary but not sufficient. We need the Hessian of the objective function to be positive definite.

$$\mathbf{H}_{\mathcal{L}} = \mathbf{H}_f + \sum_{j=1}^{p} \lambda_j \mathbf{H}_{h_j}$$

Equality Constraints

Theorem: 2^{st}-order optimality conditions

The second-order sufficient conditions are as follows:

$$\mathbf{p}^\top \mathbf{H}_{\mathcal{L}} \mathbf{p} > 0 \qquad \text{for all } \mathbf{p} \text{ such that } \mathbf{J}_{\mathbf{h}} \mathbf{p} = 0$$

This conditions assumes that the gradients of the constraints are linearly independent; that is, $\mathbf{J}_{\mathbf{h}}$ has full row rank.

Equality Constraints

Discuss and sketch the feasible region described by the equality constraints

$$-x_1 + x_3 - 1 = 0$$
$$x_1^2 + x_2^2 - 2x_1 = 0$$

The Jacobian of the constraints is given by

$$\mathbf{J}_h(\mathbf{x}) = \begin{bmatrix} -1 & 0 & 1 \\ 2x_1 - 2 & 2x_2 & 0 \end{bmatrix}$$

which has rank 2 by giving any values of x_2.

- The $\mathbf{J}_h(\mathbf{x})$ has rank less than 2 when $\mathbf{x} = \begin{bmatrix} 1 & 0 & x_3 \end{bmatrix}^\top$.
- Sine $\mathbf{x} = \begin{bmatrix} 1 & 0 & x_3 \end{bmatrix}^\top$ does not satisfy the circle constrain, any point $\mathbf{x}$ satisfying both constraints is regular. (make $\mathbf{J}_h$ has full row rank.)

Equality Constraints: Example I

Consider a constrained problem with a linear objective function and a quadratic equality constraint:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1 + 2x_2 \\ \text{subject to} \quad & h(\mathbf{x}) = \frac{1}{4}x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 + 2x_2 + \lambda\left(\frac{1}{4}x_1^2 + x_2^2 - 1\right)$$

Then,

$$\begin{aligned} \nabla\mathcal{L}_{\mathbf{x}} &= \begin{bmatrix} 1 + \frac{1}{2}\lambda x_1 \\ 2 + 2\lambda x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \nabla\mathcal{L}_{\lambda} &= \frac{1}{4}x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

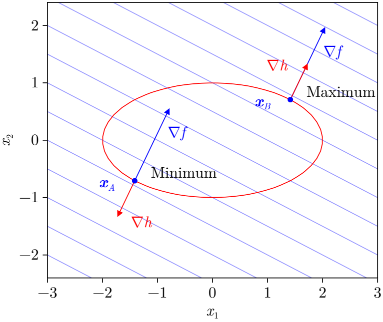
We have $x_1 = -2/\lambda$, and $x_2 = -1/\lambda$, then $\lambda = \pm\sqrt{2}$.

Equality Constraints: Example I

For each $\lambda_A = \sqrt{2}$ and $\lambda_B = -\sqrt{2}$, we obtain two possible solutions:

$$\mathbf{x}_A = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}, \quad \lambda_A = \sqrt{2}$$

$$\mathbf{x}_B = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \quad \lambda_B = -\sqrt{2}$$



- The Hessian of the Lagrangian is

$$\mathbf{H}_{\mathcal{L}} = \begin{bmatrix} \frac{1}{2}\lambda & 0 \\ 0 & 2\lambda \end{bmatrix}$$

- It is clear that $\mathbf{H}$ is positive for $\mathbf{x}_A$, and negative for $\mathbf{x}_B$. Then $\mathbf{x}_A$ is a minimum point, and $\mathbf{x}_B$ is a maximum point.

Equality Constraints: Example II

Consider the following problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1^2 + 3(x_2 - 2)^2 \\ \text{subject to} \quad & h(\mathbf{x}) = \beta x_1^2 - x_2 = 0, \end{aligned}$$

where β is a parameter that we will vary to change the characteristics of the constraint. The Lagrangian for this problem is

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \lambda) &= x_1^2 + 3(x_2 - 2)^2 + \lambda \left(\beta x_1^2 - x_2\right) \\ \nabla_{\mathbf{x}} \mathcal{L} &= \begin{bmatrix} 2x_1(1 + \lambda\beta) \\ 6(x_2 - 2) - \lambda \end{bmatrix} = 0 \\ \nabla_{\lambda} \mathcal{L} &= \beta x_1^2 - x_2 = 0 \end{aligned}$$

Form $2x_1(1 + \lambda\beta) = 0$ we get $x_1 = 0$, then the solution is $\begin{bmatrix} x_1 & x_2 & \lambda \end{bmatrix}$ $= \begin{bmatrix} 0 & 0 & -12 \end{bmatrix}$, which is independent of β.

Equality Constraints: Example II

To determine if this is a minimum, we must check the second-order conditions by evaluating the Hessian of the Lagrangian,

$$\mathbf{H}_{\mathcal{L}} = \begin{bmatrix} 2(1 - 12\beta) & 0 \\ 0 & 6 \end{bmatrix}$$

- The feasible directions are all $\mathbf{p}$ such that $\mathbf{J}_h^\top \mathbf{p} = 0$. Here $\mathbf{J}_h^\top = \begin{bmatrix} 2\beta x_1 & -1 \end{bmatrix}$, yielding $\mathbf{J}_h(\mathbf{x}^*) = \begin{bmatrix} 0 & -1 \end{bmatrix}^\top$
- The feasible directions at the solution can be represented as $\mathbf{p} = \begin{bmatrix} \alpha & 0 \end{bmatrix}^\top$, where α is any number.
- For positive curvature in the feasible directions, we require that

$$\mathbf{p}^\top \mathbf{H}_{\mathcal{L}} \mathbf{p} = 2\alpha^2(1 - 12\beta) > 0$$
$$\beta < \frac{1}{12}$$

Equality Constraints: Example III

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 + x_2 + 2 = 0 \end{aligned}$$

- Form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda h(\mathbf{x}) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 + 2)$$

$$\nabla_{\mathbf{x}} \mathcal{L} = \begin{bmatrix} 2x_1 + \lambda \\ 2x_2 + \lambda \end{bmatrix} = 0$$

$$\nabla_{\lambda} \mathcal{L} = x_1 + x_2 + 2 = 0$$

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}, \quad x_1^* = x_2^* = -1, \lambda^* = 2$$

Inequality Constraints

We can use some of the concepts from the equality constrained optimality conditions for inequality constrained problems.

- An inequality constraint j is feasible when $g_j(\mathbf{x}^*) \leq 0$ and it is said to be **active** if $g_j(\mathbf{x}^*) = 0$ and **inactive** if $g_i(\mathbf{x}^*) < 0$.
- Based on the Taylor series, for any small enough feasible step $\mathbf{p}$, we get the condition

$$f(\mathbf{x}^* + \mathbf{p}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top \mathbf{p}$$
$$\nabla f(\mathbf{x}^*)^\top \mathbf{p} \geq 0, \text{ since } \mathbf{x} \text{ is the optimal point.}$$

- The decent directions, if it is feasible, is in the open half-space defined by the hyperplane tangent to the gradient of the objective.
- Consider the Taylor series of the inequality constraints

$$g_j(\mathbf{x} + \mathbf{p}) \approx g_j(\mathbf{x}) + \nabla g_j(\mathbf{x})^\top \mathbf{p} \leq 0, \qquad j = 1, \ldots, q$$

Inequality Constraints

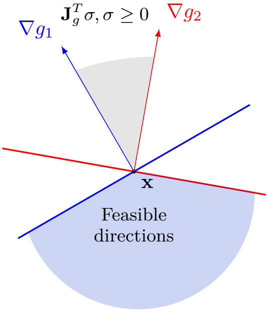
There are two possibilities to consider for each inequality constraint: inactive $g_j(\mathbf{x}) < 0$ or active $g_j(\mathbf{x}) = 0$.

- If the constraint is inactive we can take a step $\mathbf{p}$ in any direction and remain feasible as long as the step is small enough.
- Inequality constraints do not need the nullspace of the Jacobian matrix. From

$$g_j(\mathbf{x} + \mathbf{p}) \approx g_j(\mathbf{x}) + \nabla g_j(\mathbf{x})^\top \mathbf{p} \leq 0, \qquad j = 1, \ldots, q$$

  if constraint j is active ($g_j(\mathbf{x}) = 0$), then the nearby point $g_j(\mathbf{x} + \mathbf{p})$ is only feasible if $\nabla g_j(\mathbf{x})^\top \mathbf{p} \leq 0$ for all constraints j that are active. In matrix form, we can write $J_g(\mathbf{x})\mathbf{p} \leq 0$, where the Jacobian matrix includes only the gradients of the active constraints.

Inequality Constraints



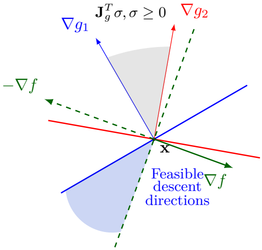
- The set of feasible directions that satisies all active constraints is the intersection of all the closed half-spaces defined by the inequality constraints, that is all $\mathbf{p}$ such that $\mathbf{J}_g(\mathbf{x})\mathbf{p} \leq 0$.
- The intersection of the feasible directions forms a polyhedral cone.
- To find the cone of feasible directions, first consider the cone formed by the active inequality constraint gradients (shown in gray).

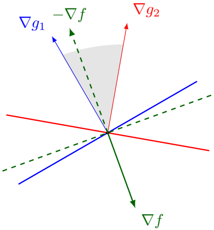
The cone is defined by all vectors $\mathbf{d}$ such that (linear combination of ∇g_j)

$$\mathbf{d} = \mathbf{J}_g^\top \sigma = \sum_{j=1}^{q} \sigma_j \nabla g_j, \quad \text{where } \sigma_j \geq 0$$

A direction $\mathbf{p}$ is feasible if $\mathbf{p}^\top \mathbf{d} \leq 0$ for all $\mathbf{d}$ in the cone. The set of all feasible directions forms the **polar cone** of the cone defined above and is shown in blue.

Inequality Constraints: Farkas' lemma





We need to establish under which condition there is no feasible descent direction or when is there no intersection between the cone of feasible directions and the open half-space of descent direction?

- There exists a $\mathbf{p}$ (dashed line) such that $\mathbf{J}_g\mathbf{p} \leq 0$ and $\nabla f^\top \mathbf{p} < 0$ (a descent direction is feasible. (above))
- There exists a σ such that $\mathbf{J}_g^\top \sigma = -\nabla f$ with $\sigma \geq 0$ (This corresponds to optimality. There is no feasible direction.(below))
- The optimality criterion for inequality constraints:

$$\nabla f + \sigma^\top \mathbf{J}_g(\mathbf{x}) = 0, \text{ with } \sigma \geq 0$$

Inequality Constraints: Farkas' lemma

- The criteria of the inequality constraints is similar to the equality constraints. However, σ corresponds to the Lagrange multipliers for the inequality constraints and carries the additional restriction that $\sigma \geq 0$ (nonnegative)
- If equality constraints are present, the conditions for the inequality constraints apply only in the subspace of the directions feasible with respect to the equality constraints.
- We can add all inequality constraints (we don't know which one we should use.) to the Lagrangian by replacing them with the equality constraint as

  $$g_j + s_j^2 = 0, \qquad j = 1, \ldots, q$$

  where s_j is a new unknown associated with each inequality constraint called a **slack variable**. This variable must be positive.
- If $s_j = 0$, the corresponding inequality constraint is active ($g_j = 0$), and when $s_j \neq 0$, the corresponding constraint is inactive.

The Lagrangian

The Lagrangian including both equality and inequality constraints is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}, \mathbf{s}) = \mathbf{f}(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{h}(\mathbf{x}) + \boldsymbol{\sigma}^\top \left(\mathbf{g}(\mathbf{x}) + \mathbf{s} \odot \mathbf{s}\right),$$

where σ represents the Lagrange multipliers associated with the inequality constraints. The $\odot$ is represented the element-wise multiplication of $\mathbf{s}$.
At the stationary point

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathcal{L} = 0 \quad &\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{l=1}^{p} \lambda_l \frac{\partial h_l}{\partial x_i} + \sum_{j=1}^{q} \sigma_j \frac{\partial g_j}{\partial x_i} = 0, i = 1, \ldots, n \\
\nabla_{\boldsymbol{\lambda}} \mathcal{L} = 0 \quad &\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial \lambda_l} = h_l = 0, \quad l = 1, \ldots, p \\
\nabla_{\boldsymbol{\sigma}} \mathcal{L} = 0 \quad &\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial \sigma_j} = g_j + s_j^2 = 0, \quad j = 1, \ldots, q \\
\nabla_{\mathbf{s}} \mathcal{L} = 0 \quad &\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial s_j} = 2\sigma_j s_j = 0, \quad j = 1, \ldots, q
\end{aligned}$$

The last one is call **complementary slackness condition**. It can help us to distinguish the active constraints from the inactive constraint.

Karush-Kuhn-Tucker (KKT) condition

Theorem: KKT 1st-order condition

$$
\begin{aligned}
\nabla \mathbf{f} + \mathbf{J}_{\mathbf{h}}^{\top} \boldsymbol{\lambda} + \mathbf{J}_{\mathbf{g}}^{\top} \boldsymbol{\sigma} &= 0 \\
\mathbf{h} &= 0 \\
\mathbf{g} + \mathbf{s} \odot \mathbf{s} &= 0 \\
\boldsymbol{\sigma} \odot \mathbf{s} &= 0 \\
\boldsymbol{\sigma} &\geq 0
\end{aligned}
$$

Theorem: 2nd-order condition

$$
\begin{aligned}
\mathbf{p}^{\top} \mathbf{H}_{\mathcal{L}} \mathbf{p} &> 0 && \text{for all } \mathbf{p} \text{ such that:} \\
\mathbf{J}_{\mathbf{h}} \mathbf{p} &= 0 \\
\mathbf{J}_{\mathbf{g}} \mathbf{p} &\leq 0 && \text{for the active constraints.}
\end{aligned}
$$

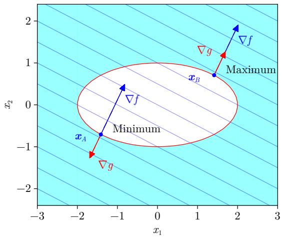
Problem with one inequality constraint

Consider a problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1 + 2x_2 \\ \text{subject to} \quad & g(\mathbf{x}) = \frac{1}{4}x_1^2 + x_2^2 - 1 \leq 0 \end{aligned}$$

The Lagrangian for this problem is

$$\mathcal{L}(x_1, x_2, \sigma, s) = x_1 + 2x_2 + \sigma\left(\frac{1}{4}x_1^2 + x_2^2 - 1 + s^2\right)$$



- Inequality constrained problem with linear objective.
- Feasible space within a ellipse.

Problem with one inequality constraint

Differentiating the Lagrangian with respect to all the variables, we get the first-order optimality conditions

$$\frac{\partial \mathcal{L}}{\partial x_1} = 1 + \frac{1}{2}\sigma x_1 = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = 2 + 2\sigma x_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \sigma} = \frac{1}{4}x_1^2 + x_2^2 - 1 = 0, \quad \frac{\partial \mathcal{L}}{\partial s} = 2\sigma s = 0$$

The last equation, we can set $s = 0$ (meaning the constraint is active) and $\sigma = 0$ (meaing the constraint is inactive). However, σ cannot be zero because the first two equation will not yield a solution. Setting that $s = 0$ and $\sigma \neq 0$, we can solve the equations to obtain:

$$\mathbf{x}_A = \begin{bmatrix} x_1 \\ x_2 \\ \sigma \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ -\frac{\sqrt{2}}{2} \\ \sqrt{2} \end{bmatrix}, \quad \mathbf{x}_B = \begin{bmatrix} x_1 \\ x_2 \\ \sigma \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ \frac{\sqrt{2}}{2} \\ -\sqrt{2} \end{bmatrix}$$

According to the KKT conditions, the Lagrange multiplier σ must be nonnegative. Point $\mathbf{x}_A$ satisfies this condition. There is no feasible descent direction a $\mathbf{x}_A$.

Problem with two inequality constraint

Consider

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1 + 2x_2 \\ \text{subject to} \quad & g_1(\mathbf{x}) = \frac{1}{4}x_1^2 + x_2^2 - 1 \le 0 \\ & g_2(\mathbf{x}) = -x_2 \le 0. \end{aligned}$$

The Lagrangian for this problem is

$$\mathcal{L}(x, \sigma, s) = x_1 + 2x_2 + \sigma_1 \left(\frac{1}{4}x_1^2 + x_2^2 - 1 + s_1^2 \right) + \sigma_2 \left(-x_2 + s_2^2 \right)$$

Differentiating the Lagrangian with respect to all the variables, we get the first-order optimality conditions

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= 1 + \frac{1}{2}\sigma_1 x_1 = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = 2 + 2\sigma_1 x_2 - \sigma_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \sigma_1} &= \frac{1}{4}x_1^2 + x_2^2 - 1 + s_1^2 = 0, \quad \frac{\partial \mathcal{L}}{\partial \sigma_2} = -x_2 + s_2^2 = 0 \\ \frac{\partial \mathcal{L}}{\partial s_1} &= 2\sigma_1 s_1 = 0, \quad \frac{\partial \mathcal{L}}{\partial s_2} = 2\sigma_2 s_2 = 0 \end{aligned}$$

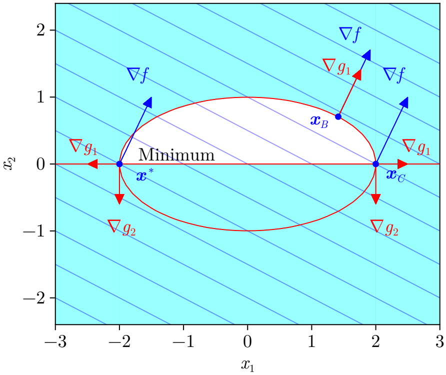
Problem with two inequality constraint

We have two complementary slackness conditions, which yield the four potential combinations listed below:

Assumption	Meaning	x_1	x_2	σ_1	σ_2	s_1	s_2	Point
$s_1 = 0$	g_1 is active	-2	0	1	2	0	0	$\mathbf{x}^*$
$s_2 = 0$	g_2 is active	2	0	-1	2	0	0	$\mathbf{x}_C$
$\sigma_1 = 0$ $\sigma_2 = 0$	g_1 is inactive g_2 is inactive	-	-	-	-	-	-	
$s_1 = 0$ $\sigma_2 = 0$	g_1 is active g_2 is inactive	$\sqrt{2}$	$\frac{\sqrt{2}}{2}$	$-\sqrt{2}$	0	0	$2^{-\frac{1}{4}}$	$\mathbf{x}_B$
$\sigma_1 = 0$ $s_2 = 0$	g_1 is inactive g_2 is active	-	-	-	-	-	-	

Assuming that both constraints are active yields two possible solutions ($\mathbf{x}^*$ and $\mathbf{x}_C$) cooresponding to two different Lagrange multipliers. According to the KKT conditions, the Lagrange multipliers for all active inequality constraints have to be positive, so only the solution with $\sigma_1 = 1$, then $\mathbf{x}^*$ is a candidate for a minimum.

The feasible region is the top half of the ellipse, as show below



Simple Constrained Example

Consider:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1^2 + x_1 x_2 + x_2^2 \\ \text{subject to} \quad & x_2 \geq 1 \\ & x_1 + x_2 \geq 3 \end{aligned}$$

Form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \sigma) = x_1^2 + x_1 x_2 + x_2^2 + \sigma_1(1 - x_2) + \sigma_2(x_1 + x_2 - 3)$$

Form necessary conditions:

$$\nabla_{\mathbf{x}} \mathcal{L} = \begin{bmatrix} 2x_1 + x_2 + \sigma_2 \\ x_1 + 2x_2 - \sigma_1 + \sigma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla_{\sigma} \mathcal{L} = \begin{bmatrix} 1 - x_2 \\ x_1 + x_2 - 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Simple Constrained Example

Consider the various options:

- $\sigma_1 = \sigma_2 = 0$, both constraints are inactive

$$\nabla_{\mathbf{x}}\mathcal{L} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\Rightarrow\; x_1 = x_2 = 0$$

- $\sigma_1 = 0$ (inactive), $\sigma_2 \geq 0$ (active)

$$\nabla_{\mathbf{x}}\mathcal{L} = \begin{bmatrix} 2x_1 + x_2 + \sigma_2 \\ x_1 + 2x_2 + \sigma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla_{\sigma_2}\mathcal{L} = x_1 + x_2 - 3 = 0$$

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \;\Rightarrow\; x_1 = x_2 = \frac{3}{2}, \sigma_2 = -\frac{9}{2}$$

Satisfy the constraints but σ_2 is negative, and $f(\mathbf{x}) = 6.75$ is not minimum.

Simple Constrained Example

- $\sigma_1 \geq 0$ (active), $\sigma_2 = 0$ (inactive)

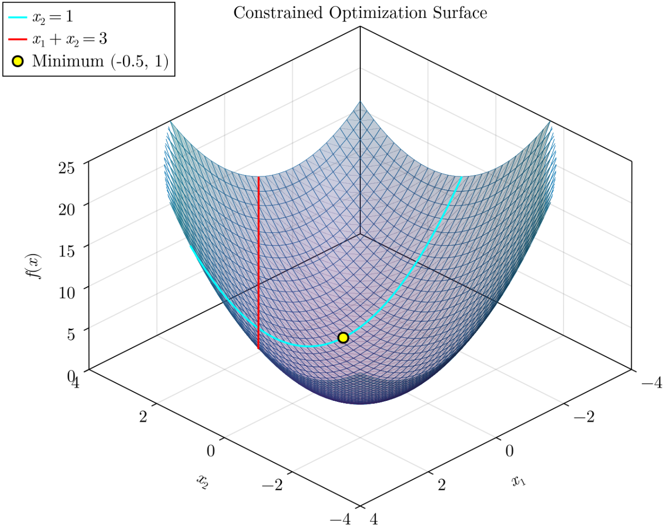
$$\nabla_{\mathbf{x}} \mathcal{L} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 - \sigma_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla_{\sigma_1} \mathcal{L} = 1 - x_2 = 0$$

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \Rightarrow x_1 = -0.5, x_2 = 1, \sigma_1 = \frac{3}{2}$$

Satisfy the constraints, σ_1 is positive, and $f(\mathbf{x}) = 0.75$ is minimum.

Simple Constrained Example



Reference

1. Joaquim R. R. A. Martins, Andrew Ning, ”*Engineering Design Optimization*,” Cambridge University Press, 2021.
2. Mykel J. kochenderfer, and Tim A. Wheeler, ”*Algorithms for Optimization*,” The MIT Press, 2019.
3. Ashok D. Belegundu, Tirupathi R. Chandrupatla, ”Optimization Concepts and Applications in Engineering,” Cambridge University Press, 2019.
4. Stephen Boyd, and Lieven Vandenberghe , ”*Convex Optimization*,” Cambridge University Press, 2009.