Lecture 7 Review Notes

MA8.401 Topics in Applied Optimization
Monsoon 2023

Contributors:
- Sriteja Reddy Pashya (2021111019)
- Romica Raisinghani (2021101053)

1 Transformation of objective and constraint function

- Suppose $\psi_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing
- $\psi_1, ..., \psi_m : \mathbb{R} \to \mathbb{R}$ satisfy:

$$\psi_i(u) \leq 0, \text{if and only if} \quad u \leq 0$$

- $\psi_{m+1}, ..., \psi_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy

$$\psi_i(u) = 0, \text{if and only if} \quad u = 0$$

- Define $\overline{f}_i$ as

$$\overline{f}_i(x) = \psi_i(f_i(x)), \quad i = 0, ..., m$$

- Define $\overline{h}_i$ as

$$\overline{h}_i(x) = \psi_{m+i}(h_i(x)), \quad i = 1, ..., p$$

The transformed model is:

$$\begin{array}{rl} \text{minimize} & \overline{f}_0(x) \\ \text{subject to} & \overline{f}_i(x) \leq 0, \quad i = 1, \ldots, m \\ & \overline{h}_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

1.1 Examples:

- **Linearization:** One common transformation is to linearize a non-linear objective function or constraint. For example, consider the non-linear objective function $f(x) = x^2$. This can be linearized by approximating it with a first-order Taylor series expansion:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

where x_0 is a reference point. For example, if we choose $x_0 = 0$, we get the following linear approximation:

$$f(x) \approx x$$

This approximation can then be used to replace the non-linear objective function in the optimization problem.

- **Logarithm transformation:** Another common transformation is to use a logarithm transformation. For example, consider the non-linear constraint $g(x) \leq b$, where $g(x)$ is a monotonically increasing function. This constraint can be transformed into a linear constraint by taking the logarithm of both sides:

$$log(g(x)) \leq log(b)$$

This transformation is often useful because it can make it easier to solve the optimization problem.

1.2 Results:

- **Equivalence of transformed problems:** It is important to note that the transformation of objective and constraint functions should not change the optimal solution to the optimization problem. In other words, the transformed problem should be equivalent to the original problem. This can be shown mathematically using the following theorem:

Theorem: Let $f(x)$ be an objective function and $g(x) \leq b$ be a constraint function. Let $h(x)$ be a monotonically increasing function and $u(x)$ be a function such that $u(x) \leq 0$ if and only if $x < 0$. Then the following two optimization problems are equivalent:

$$\begin{aligned} &\text{minimize} f(x) \\ &\text{subject to} g(x) \leq b \end{aligned}$$

and,

$$\begin{aligned} &\text{minimize} \quad h(f(x)) \\ &\text{subject to} \quad u(g(x)) \leq u(b) \end{aligned}$$

- **Existence of optimal solutions:** It is also important to note that the transformation of objective and constraint functions should not make the optimization problem infeasible. In other words, the transformed problem should still have an optimal solution. This can be shown mathematically using the following theorem:

Theorem: If the original optimization problem has an optimal solution, then the transformed optimization problem also has an optimal solution.

1.3 Illustrations:

- **Linearization example:** Consider the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad x^2 \\ &\text{subject to} \quad x \geq 0 \end{aligned}$$

This problem can be solved by linearizing the objective function as follows:

$$\begin{aligned} &\text{minimize} \quad x \\ &\text{subject to} \quad x \geq 0 \end{aligned}$$

The optimal solution to this problem is $x = 0$, which is also the optimal solution to the original problem.

- **Logarithm transformation example:** Consider the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad e^x \\ &\text{subject to} \quad e^x \leq 10 \end{aligned}$$

This problem can be solved by transforming the constraint function using a logarithm transformation as follows:

$$\begin{aligned} &\text{minimize} \quad e^x \\ &\text{subject to} \quad x \leq ln(10) \end{aligned}$$

The optimal solution to this problem is $x = ln(10)$, which is also the optimal solution to the original problem.

2 Slack variables

Given the optimization problem in standard form:

Optimization Problem(Standard form):

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

An observation to make is that we can replace inequality constraints by equality constraints and non-negativity constraints i.e $f_i(x) \leq 0$ if and only if there is an $s_i \geq 0$ that satisfies $f_i(x) + s_i = 0$. Using this transformation we obtain the problem:

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{subject to} & s_i \geq 0, & i = 1, \dots, m \\ & f_i(x) + s_i = 0, & i = 1, \dots, m \\ & h_i(x) = 0, & i = 1, \dots, p \end{array}$$

where the variables are $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$

This problem has $n + m$ variables, m inequality constraints (the nonnegativity constraints on s_i), and $m + p$ equality constraints.
The new variable s_i is called the **slack variable** associated with the original inequality constraint $f_i(x) \leq 0$. Introducing slack variables replaces each inequality constraint with an equality constraint, and a non-negativity constraint.

2.1 Example:

Consider the following linear programming problem in standard form:

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

where $c \in R^n$, $A \in R^{m \times n}$, and $b \in R^m$. This problem can be transformed into the following equivalent problem using slack variables:

$$\begin{array}{rll} \text{minimize} & c^T x & \\ \text{subject to} & Ax + s = b & s \geq 0 \end{array}$$

where $s \in R^m$ are the slack variables.

2.2 Results:

The following theorem provides a necessary and sufficient condition for the feasibility of the linear programming problem in standard form:

Theorem: The linear programming problem in standard form is feasible if and only if the following system of inequalities is feasible:

$$Ax + s = b \qquad s \geq 0$$

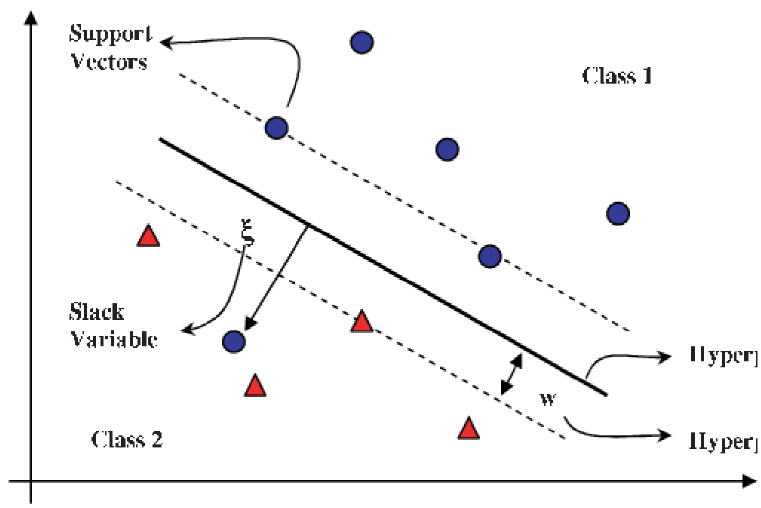
This theorem shows that the slack variables can be used to check the feasibility of the linear programming problem.

2.3 Illustrations:

An illustration of slack variables is in the context of support vector machines (SVMs). SVMs are a type of machine learning algorithm used for classification and regression.

In the context of SVMs, slack variables are used to allow some of the training data points to fall outside of the margin. This is necessary because real-world data is often noisy and cannot be perfectly classified or fitted by a linear model.

The following diagram illustrates the use of slack variables in SVMs: The blue line represents the



decision boundary. The green circles represent the support vectors. The red circles represent the training data points that fall outside of the margin.

The slack variables are used to measure the distance between the training data points and the decision boundary. The goal of the SVM algorithm is to find the decision boundary that maximizes the margin and minimizes the sum of the slack variables.

Here is a more concrete example:

Suppose we are training an SVM to classify emails as spam or ham. We have a set of training data points, each of which is represented by a vector of features, such as the number of words in the email, the number of punctuation marks, and the presence of certain keywords.

We want to train the SVM to find a decision boundary that separates the spam emails from the ham emails. However, some of the emails in the training set may be misclassified, even by the best possible decision boundary.

To allow for this, we use slack variables. The slack variable for each training data point measures how far the data point falls from the decision boundary. The goal of the SVM algorithm is to find the decision boundary that maximizes the margin and minimizes the sum of the slack variables.

This means that the SVM algorithm will try to find a decision boundary that separates the spam emails from the ham emails as widely as possible, while also minimizing the number of misclassified emails.

Slack variables are a powerful tool that can be used to improve the performance of machine learning algorithms. They are also a useful concept to understand in the context of optimization problems.

3 Convex Optimization Problem

A convex optimization problem is one of the form:

Convex Optimization Problem(Standard form):

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & {a_i}^T x = b_i, & i = 1, \ldots, p \end{array}$$

where $f_0, ..., f_m$ are convex functions.

Comparing the convex optimization problem with the general standard form problem, the convex problem has three additional requirements:

- the objective function must be convex,
- the inequality constraint functions must be convex,
- the equality constraint functions $h_i(x) = {a_i}^T x - b_i$ must be affine.

The feasible set of a convex optimization problem is convex, since it is the intersection of the domain of the problem

$$D = \bigcap_{i=0}^{m} \mathrm{dom} f_i$$

which is a convex set, with m (convex) sublevel sets $\{x | f_i(x) \leq 0\}$ and p hyperplanes $\{x | {a_i}^T x = b_i\}$.

If f_0 is quasiconvex instead of convex, we say the above problem is a (standard form) **quasiconvex optimization problem**. Since the sublevel sets of a convex or quasiconvex function are convex, we conclude that for a convex or quasiconvex optimization problem the ϵ-suboptimal sets are convex. In particular, the optimal set is convex. If the objective is strictly convex, then the optimal set contains at most one point.

3.1 Examples:

Consider the following optimization problem with $x \in \mathbb{R}^2$

The transformed model is:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = {x_1}^2 + {x_2}^2 \\ \text{subject to} \quad & f_1(x) = \frac{x_1}{1 + {x_2}^2} \leq 0, \\ & h_1(x) = (x_1 + x_2)^2 = 0, \end{aligned}$$

a) Is this problem a convex optimization problem?
b) Can you rewrite this in convex optimization problem?

Solution:
a) This problem is **not a convex optimization problem** in standard form since the equality constraint function h_1 is not affine, and the inequality constraint function f_1 is not convex. Nevertheless the feasible set, which is $\{x | x_1 \leq 0, x_1 + x_2 = 0\}$, is convex. So although in this problem we are minimizing a convex function f_0 over a convex set, it is not a convex optimization problem by our definition.

b) The problem is readily reformulated as:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = {x_1}^2 + {x_2}^2 \\ \text{subject to} \quad & \overline{f_1}(x) = x_1 \leq 0, \\ & \overline{h_1}(x) = x_1 + x_2 = 0, \end{aligned}$$

which is in standard convex optimization form, since f_0 and $\overline{f_1}$ are convex, and $\overline{h_1}$ is affine.

3.2 Results:

- **Existence of optimal solution:** Convex optimization problems always have an optimal solution. This is because the feasible region of a convex optimization problem is always convex, and the objective function is always convex or quasiconvex.
- **Uniqueness of optimal solution:** Convex optimization problems can have a unique optimal solution or multiple optimal solutions. The uniqueness of the optimal solution depends on the specific problem.

3.3 Illustrations:

Problem: Imagine you are a farmer, and you have a rectangular piece of land that you want to fence in. You have a fixed amount of fencing material, let's say 100 meters of fencing. You want to maximize the area of the enclosed rectangular field.

Variables: L - The length of the rectangular field.
W- The width of the rectangular field.

Objective: Maximize the area A of the rectangular field, which is given by $A = L \cdot W$

Constraints:

- The total amount of fencing material used should not exceed 100 meters: $2L + 2W \leq 100$.
- The length and width of the field should be non-negative: $L \geq 0$ and $W \geq 0$.

This problem can be formulated as a convex optimization problem. We want to find the values of L and W that maximize the area A while satisfying the constraints.
Mathematically, the convex optimization problem can be written as:

$$
\begin{aligned}
&\text{Maximize} \quad A = L \cdot W \\
&\text{subject to:} \qquad\qquad\qquad\qquad\qquad 2L + 2W \leq 100 \\
&\qquad\qquad\quad L \geq 0 \\
&\qquad\qquad\quad W \geq 0
\end{aligned}
$$

The objective function (maximizing area) is a convex function, and the constraints are linear, making this a convex optimization problem. Solving this problem will give you the dimensions of the rectangular field that maximize the enclosed area while using no more than 100 meters of fencing material.

4 Concave maximization problem

With a slight abuse of notation, we will also refer to:

Concave Optimization Problem(Standard form):

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & {a_i}^T x = b_i, \quad i = 1, \ldots, p \end{array}$$

where $f_0, ..., f_m$ are convex functions.

as a convex optimization problem if the objective function f_0 is concave, and the inequality constraint functions $f_1, ..., f_m$ are convex. This concave maximization problem is readily solved by minimizing the convex objective function $-f_0$.

4.1 Examples:

- **Portfolio Optimization:** In finance, the goal is to maximize the expected return of a portfolio while considering the risk. The expected return can be formulated as a concave function, and constraints may include budget constraints or limits on the risk exposure.
- **Production Planning:** In manufacturing, a company may want to maximize its profit by choosing the optimal production quantities for various products. Profit can be considered a concave function, and constraints may include resource availability or production capacity.

4.2 Results:

- **First-Order Condition for Optimality (Karush-Kuhn-Tucker conditions):** In concave maximization, the first-order condition for optimality is:

  $$f'(x) + \sum_{i=1}^{m} \lambda_i \cdot g_i'(x) = 0$$

  where x is the optimal solution, $f'(x)$ is the derivative of the objective function, and $g_i'(x^)$ is the derivative of the constraint functions. λ_i are the Lagrange multipliers.
- **Second-Order Condition for Concavity:** In concave maximization, if the objective function $f(x)$ is twice continuously differentiable, and the constraints are convex, a necessary and sufficient condition for global optimality is that the Hessian matrix of the Lagrangian is negative semi-definite.

4.3 Illustration:

Consider a simple concave maximization problem: maximizing the concave function $f(x) = -x^2$ subject to the constraint $x \leq 1$.

- The objective function is concave as it has a downward-facing parabolic shape.
- The constraint is $x \leq 1$, which can be visualized as a horizontal line at $x = 1$.

  To illustrate this, you can create a graph with the x and $f(x)$ axes. Draw the concave curve of $f(x) = -x^2$, and indicate the constraint line $x = 1$. The optimal solution will be where the concave curve touches the constraint line, which in this case is $x = 1$, and the maximum value of $f(x)$ is 0.

  This simple illustration demonstrates the concept of concave maximization, where you maximize a concave function subject to constraints.

5 Optimality Criteria for Constraint Optimization Problem

Suppose that the objective f_0 in a convex optimization problem is differentiable, so that for all $x, y \in dom f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (y - x) \tag{1}$$

Let X denote the feasible set, i.e.,

$$X = \{x \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\} \tag{2}$$

Then x is optimal if and only if $x \in X$ and

$$\nabla f_0(x)^T (y - x) \geq 0 \quad \text{for all} \quad y \in X \tag{3}$$

This optimality criterion can be understood geometrically: If $\nabla f_0(x) \neq 0$, it means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x (see figure below). Optimality criteria are

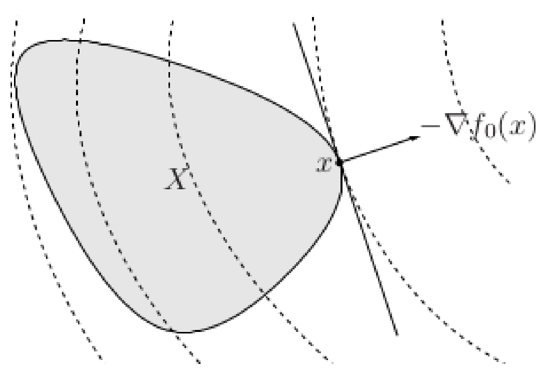


Figure 1: Geometric interpretation of the optimality condition. The feasible set X is shown shaded. Some level curves of f_0 are shown as dashed lines. The point x is optimal: $-\nabla f_0(x)$ defines a supporting hyperplane (shown as a solid line) to X at x.

essential for solving constraint optimization problems. These criteria help identify optimal solutions without directly solving for them. Let's discuss optimality criteria for constraint optimization problems, provide examples, results, and an illustration.

Optimality Criteria:

- **Kuhn-Tucker (KT) Conditions:** Also known as Karush-Kuhn-Tucker conditions, these are the general optimality conditions for constraint optimization. For a maximization problem, the Kuhn-Tucker conditions are:

  - **Stationarity:** $\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0$, where x is the optimal solution, λ_i^* are the Lagrange multipliers, and ∇ represents the gradient.

  - **Primal feasibility:** $g_i(x^*) \leq 0$ for all i (constraint satisfaction).

  - **Dual feasibility:** $\lambda_i^* \geq 0$ for all i (non-negativity of multipliers).

  - **Complementary slackness:** $\lambda_i^* g_i(x) = 0$ for all i (either constraint is active, $\lambda_i > 0$, or it does not affect the objective, $g_i(x^*) \leq 0$).

5.1 Examples:

- **Linear Programming (LP):** Consider a linear programming problem where you want to maximize profit while subject to linear constraints (e.g., resource constraints, budget constraints). The KT conditions are applied to determine the optimal solution, including Lagrange multipliers.
- **Support Vector Machines (SVM):** In machine learning, SVM aims to find the optimal hyperplane for classification. The KT conditions help identify the optimal hyperplane and support vectors by considering margin constraints and hinge loss.

5.2 Results:

- **Strong Duality:** In convex optimization, if Slater's condition holds, strong duality is achieved. This means that the optimal objective value of the dual problem is equal to the optimal objective value of the primal problem.
- **Sensitivity Analysis:** Optimality criteria are useful for sensitivity analysis, which assesses how changes in problem parameters or constraints affect the optimal solution and objective value

5.3 Illustration:

Consider a simple linear programming problem where you want to maximize profit while subject to budget and resource constraints. The objective function is to maximize profit (P), and the constraints are as follows:

- $2x + 3y \leq 12$ (Resource constraint)
- $4x + 2y \leq 8$ (Budget constraint)

You can illustrate this on a graph, with x and y on the axes. Plot the lines representing the constraints, and shade the feasible region. The optimal solution can be found using the KT conditions and Lagrange multipliers.
In this illustration, you can show how the KT conditions help determine the optimal values of x and y, the Lagrange multipliers, and the maximum profit.

6 Convex Optimization Problem: Local Optima = Global Optima

FACT: For a convex optimization problem, any local optima is a global optima

Proof: Let x^* be a local minimizer of f_0 on the set X, and let $y \in X$. By definition, $x^* \in dom f_0$
We need to prove that $f_0(y) \geq f_0(x^*) = p^*$. There is nothing to prove if $f_0(y) = +\infty$, so let us assume that $y \in dom f_0$. By convexity of f_0 and X, we have $x_\theta := \theta y + (1 - \theta)x^* \in X$, and:

$$f_0(x_\theta) - f_0(x^*) \leq \theta(f_0(y) - f_0(x^*))$$

Since x^* is a local minimizer, the left hand side in this inequality is nonnegative for all small enough values of $\theta > 0$. We conclude that the right hand side is nonnegative, i.e., $f_0(y) \geq f_0(x^*)$, as claimed.
Also, the optimal set is convex, since it can be written

$$X^{\text{opt}} = x \in \mathbb{R}^n : f_0(x) \leq p^*, x \in X$$

This ends our proof.

6.1 Examples:

- **Linear Programming (LP):** Linear programming problems are convex optimization problems. If you find a local optimum for a linear programming problem, it will always be the global optimum. For example, consider a problem of maximizing profit subject to linear constraints on resources and budgets.
- **Quadratic Programming (QP):** Quadratic programming problems are also convex. If you find a local minimum for a quadratic programming problem, it will be the global minimum. An example is minimizing the cost of a quadratic function subject to linear constraints.

6.2 Results:

- **Convex Functions:** The key to this property lies in the convexity of the objective function and the convexity of the feasible set defined by the constraints. Convex functions have the property that any local minimum (maximum) is a global minimum (maximum).

- **First-Order Condition for Convexity:** A function $f(x)$ is convex if and only if its first derivative is monotonically increasing. This means that if you find a local minimum (or maximum), it must be a global minimum (or maximum) since the function is always increasing (decreasing) in that neighborhood.
- **Geometric Interpretation:** Geometrically, for a convex optimization problem, the feasible set forms a convex set, and the objective function is a convex function. This means that any line segment connecting two points in the feasible set lies entirely within the feasible set, and the function lies below the chord connecting the function values at those points.

6.3 Illustration:

Consider a simple convex optimization problem where you want to minimize a convex quadratic function subject to linear constraints:

$$\begin{aligned} &\text{minimize} \quad f(x) = x^2 \\ &\text{subject to} \quad x \geq 0 \end{aligned}$$

For this problem, the objective function $f(x)$ is a convex quadratic function, and the constraint $x \geq 0$ forms a convex set. In the illustration, you can visualize the convex function as an upward-facing parabola and the constraint as the non-negative region of the x-axis.

Now, consider a local minimum within the feasible set (e.g., $x = 2$). You will notice that the function is monotonically increasing around this local minimum, and the minimum you found locally is indeed the global minimum.

This simple example illustrates the property that for convex optimization problems, any local optimum is also the global optimum due to the convexity of the objective function and the convexity of the feasible set.

7 Convex Optimization Problem: Optimality Criteria

FACT: If f_0 in a convex optimization problem is differentiable, then the point x is optimal if $\nabla f_0(x)^T(y-x) \geq 0$ for all $y \in X$

Proof: First suppose $x \in X$ and satisfies (3). Then if $y \in X$ we have, by (1), $f_0(y) \geq f_0(x)$. This shows x is an optimal point for (2).
Conversely, suppose x is optimal, but the condition (3) does not hold, i.e., for some $y \in X$ we have

$$\nabla f_0(x)^T(y-x) < 0$$

Consider the point $z(t) = ty + (1-t)x$, where $t \in [0,1]$ is a parameter. Since $z(t)$ is on the line segment between x and y, and the feasible set is convex, $z(t)$ is feasible. We claim that for small positive t we have $f_0(z(t)) < f_0(x)$, which will prove that x is not optimal. To show this, note that

$$\frac{d}{dt} f_0(z(t)) \bigg|_{t=0} = \nabla f_0(x)^T (y - x) < 0$$

so for small positive t, we have $f_0(z(t)) < f_0(x)$. This ends our proof.

7.1 Examples:

- **Linear Programming (LP):** In linear programming, you aim to optimize a linear objective function subject to linear constraints. If the objective function is convex and differentiable, you can use the first-order optimality condition to check for optimality.
- **Support Vector Machines (SVM):** In SVM, the objective function is convex and differentiable. You can apply the first-order optimality condition to find the optimal hyperplane for classification.

7.2 Results:

First-Order Condition for Convexity: The condition is a direct consequence of the convexity of the objective function $f_0(x)$. A convex function has the property that any linear approximation to it is a global underestimate. The gradient of the function $\nabla f_0(x)$ is the direction of steepest ascent, and for the point to be optimal, the dot product of the gradient and the difference $y - x$ must be non-negative for all y in the feasible set.

7.3 Illustration:

Consider a simple convex optimization problem with a one-dimensional convex objective function $f_0(x)$, which is a convex parabola. The feasible set X consists of all x values.
1. Plot the convex objective function $f_0(x)$ as a parabolic curve, with the minimum point being the optimal solution.
2. Choose a point x within the feasible set.
3. Calculate the gradient at point x, which gives the direction of steepest ascent.
4. Draw a vector from x to y, representing the difference $y - x$, where y is another point in the feasible set.
5. Calculate the dot product $\nabla f_0(x)^T (y - x)$ to check if it's non-negative. If it is, then x satisfies the first-order optimality condition.

This illustration visually shows how the condition ensures that for a convex optimization problem with a differentiable convex objective function, the gradient at the optimal solution points in a direction that is non-negative for all y in the feasible set. This is a key characteristic of convex optimization problems.

8 Convex Optimization Problem: Optimality for Unconstrained problems

Fact: For an unconstrained problem, $(m = p = 0,)$, the optimality condition

$$\nabla f_0(x)^T(y - x) \geq 0$$

reduces to the well known necessary and sufficient condition

$$\nabla f_0(x) = 0$$

for x to be optimal.

Proof:
Suppose x is optimal, which means here that $x \in dom f_0$, and for all feasible y we have $\nabla f_0(x)^T(y - x) \geq 0$. Since f_0 is differentiable, its domain is (by definition) open, so all y sufficiently close to x are feasible. Let us take $y = x - t\nabla f_0(x)$, where $t \in \mathbb{R}$ is a parameter. For t small and positive, y is feasible, and so

$$\nabla f_0(x)^T(y - x) = -t\|\nabla f_0(x)\|_2^2 \geq 0$$

from which we conclude $\nabla f_0(x) = 0$.
This ends our proof.

8.1 Examples:

- **Quadratic Optimization:** Consider a quadratic optimization problem where you aim to minimize a convex quadratic function $f_0(x)$ without any constraints. This is a typical unconstrained convex optimization problem where $f_0(x) = 0$ characterizes the optimal solution.
- **Convex Optimization in Machine Learning:** In various machine learning algorithms, unconstrained convex optimization problems arise, such as in training linear regression models or support vector machines. The condition $f_0(x) = 0$ ensures the optimal solution.

8.2 Results:

Necessary and Sufficient Condition: When there are no constraints $(m = p = 0)$, the gradient $f_0(x) = 0$ is both necessary and sufficient for a point to be optimal in an unconstrained convex optimization problem.

8.3 Illustration:

Consider a simple convex quadratic optimization problem with an objective function $f_0(x) = x^2$ This problem has no constraints.

1. Plot the convex quadratic function $f_0(x) = x^2$, which is a parabolic curve.
2. The optimal solution for this unconstrained problem can be found by solving for $\nabla f_0(x) = 0$, which results in $2x = 0$ and $x = 0$.
3. To illustrate the optimality condition, draw the gradient vector $\nabla f_0(x)$ at $x = 0$. It will be a zero vector.
4. Take any point y on the curve. Calculate the dot product $\nabla f_0(x)^T(y-x)$. It will be non-negative for all y, demonstrating that $\nabla f_0(x) = 0$ is the necessary and sufficient condition for optimality.

This simple illustration visually demonstrates that for unconstrained convex optimization problems, the condition $\nabla f_0(x) = 0$ characterizes the optimal solution. In the context of this problem, the solution is $x = 0$, and the gradient at that point is zero, indicating optimality.

9 Unconstrained Quadratic Optimization

Consider the problem of minimizing the quadratic function

$$f_0(x) = \frac{1}{2}x^T Px + q^T x + r,$$

where $P \in \mathbb{S}^n_+$ (which makes f_0 convex). The necessary and sufficient condition for x to be a minimizer of f_0 is

$$\nabla f_0(x) = Px + q = 0.$$

Several cases can occur, depending on whether this (linear) equation has no solutions, one solution, or many solutions.

- If $q \notin \mathcal{R}(P)$, then there is no solution. In this case f_0 is unbounded below.
- If $P > 0$ (which is the condition for f_0 to be strictly convex), then there is a unique minimizer, $x^* = -P^{-1}q$.
- If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is the (affine) set $X_{opt} = -P^\dagger q + \mathcal{N}(P)$, where $P^\dagger$ denotes the pseudo-inverse of P and $\mathcal{N}(P)$ is the null space of P.

9.1 Examples:

- **Portfolio Optimization:** In finance, portfolio optimization problems involve finding an optimal allocation of assets to minimize risk or maximize return. These problems can often be formulated as unconstrained quadratic optimization problems.

- **Machine Learning:** In machine learning, support vector machines (SVM) optimization involves minimizing a quadratic cost function to find the optimal hyperplane for classification.

9.2 Results:

- **Convexity of f_0:** The convexity of f_0 depends on the positive definiteness of matrix P. When P is positive definite, f_0 is strictly convex, and there is a unique minimizer.
- **Unboundedness:** If q is not in the range of P, the problem is unbounded below, meaning there is no minimum.
- **Pseudo-Inverse:** In the case of a singular P with q in the range of P, the pseudo-inverse $P^\dagger$ and the null space $\mathcal{N}(P)$ are involved in defining the set of optimal points.

9.3 Illustration:

Consider a simple unconstrained quadratic optimization problem with P being a positive definite matrix. Illustrate the following:
1. Plot the quadratic function $f_0(x)$ as a convex parabolic curve.
2. Show the unique minimizer $x^* = -P^{-1}q$ as the point where the gradient of $f_0(x)$ is zero, i.e., $Px^* + q = 0$.
3. Demonstrate that this minimizer is indeed the global minimum of the function.

This illustration visually represents the concept of unconstrained quadratic optimization, showing the unique minimizer and the convex nature of the objective function.

10 Problems with Equality Constraints

Consider the following optimization problem with equality constraints only

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b \end{array}$$

Feasible set is affine, recall that x is feasible if it satisfies

$$\nabla f_0(x)^T(y - x) \geq 0, \quad \text{for all} \quad y \quad \text{such that} \quad Ay = b$$

Since x is feasible, every feasible y has the form $y = x + v$ for some $v \in \mathcal{N}(A)$. The optimality condition can therefore be expressed as:

$$\nabla f_0(x)^T v \geq 0 \quad \text{for all} v \in \mathcal{N}(A)$$

If a linear function is nonnegative on a subspace, then it must be zero on the subspace, so it follows that $\nabla f_0(x)^T v = 0$ for all $v \in \mathcal{N}(A)$. In other words,

$$\nabla f_0(x) \perp \mathcal{N}(A)$$

Using the fact that $\mathcal{N}(A) \perp = \mathcal{R}(A^T)$, this optimality condition can be expressed as $\nabla f_0(x) \in \mathcal{R}(A^T)$, i.e., there exists a $v \in R^p$ such that

$$\nabla f_0(x) + A^T v = 0.$$

Together with the requirement $Ax = b$(i.e., that x is feasible), this is the classical Lagrange multiplier optimality condition.

10.1 Examples:

- **Portfolio optimization:** An investor wants to maximize their expected return on investment while minimizing their risk. This can be formulated as an optimization problem with equality constraints, where the equality constraints represent the investor's budget and risk tolerance.
- **Production planning:** A company wants to minimize the cost of producing a certain number of units of a product. This can be formulated as an optimization problem with equality constraints, where the equality constraints represent the company's production requirements.
- **Transportation planning:** A company wants to minimize the cost of transporting goods from one location to another. This can be formulated as an optimization problem with equality constraints, where the equality constraints represent the company's transportation requirements.

10.2 Results:

- The feasible set is affine. This means that if two points are feasible, then any line segment connecting them is also feasible.
- The optimality condition can be expressed as:

  $$Vfo(z) \geq 0 \text{ for all } u \in N(A)$$

  where Vfo is the gradient of f_o and $N(A)$ is the null space of A.
- The Lagrange multiplier optimality condition is:

  $$Vfo(z) + Av = 0$$

  for some $v \in \mathbb{R}^P$, where P is the column space of A.

10.3 Illustration:

- Consider the following portfolio optimization problem:

$$\min_{\sigma^2} \sigma^2$$

subject to

$$\mathbb{E}[R] = \mu$$

where σ^2 is the variance of the portfolio return and $\mathbb{E}[R]$ is the expected return of the portfolio. The equality constraint represents the investor's expected return requirement.

To solve this problem, we can use the Lagrange multiplier method. The Lagrange function is:

$$L(\sigma^2, \mu, \lambda) = \sigma^2 + \lambda(\mathbb{E}[R] - \mu)$$

where λ is the Lagrange multiplier.

The optimality condition is:

$$\nabla L(\sigma^2, \mu, \lambda) = 0$$

which gives us the following system of equations:

$$2\sigma^2 = \lambda$$
$$\mathbb{E}[R] = \mu$$

Solving this system of equations, we get the optimal solution:

$$\sigma^2 = \frac{\lambda}{2}$$
$$\mu = \mathbb{E}[R]$$

- Consider the following production planning problem:

$$\min_{x} c(x)$$

subject to

$$Ax = b$$

where $c(x)$ is the cost of producing x units of the product, A is a matrix of production coefficients, and b is a vector of production requirements. The equality constraint represents the company's production requirements.

To solve this problem, we can use the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions are:

$$\nabla c(x) + A^T \lambda = 0$$

$$Ax = b$$

$$\lambda \geq 0$$

$$\lambda^T(Ax - b) = 0$$

where λ is a vector of Lagrange multipliers.

Solving the KKT conditions, we get the optimal solution:

$$x = A^{-1}b$$

$$\lambda = -(\nabla c(x))^T A^{-1}$$

11 Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Concave Optimization Problem(Standard form):

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & {a_i}^T x = b_i, & i = 1, \ldots, p \end{array}$$

where $f_0, ..., f_m$ are convex functions.

For a convex problem the equality constraints must be linear, i.e., of the form $Ax = b$. In this case they can be eliminated by finding a particular solution x_0 of $Ax = b$, and a matrix F whose range is the nullspace of A, which results in the problem

$$\begin{array}{rll} \text{minimize} & f_0(Fz + x_0) & \\ \text{subject to} & f_i(Fz + x_0) \leq 0, & i = 1, \ldots, m \end{array}$$

with variable z. Since the composition of a convex function with an affine function is convex, eliminating equality constraints preserves convexity of a problem.
Moreover, the process of eliminating equality constraints (and reconstructing the solution of the original problem from the solution of the transformed problem) involves standard linear algebra operations.

11.1 Examples:

- **Example 1:** Consider the following convex optimization problem in standard form:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) \leq 0, \qquad\qquad\qquad Ax = b \end{aligned}$$

  where f and h are convex functions, and A is a matrix. We can eliminate the equality constraint $Ax = b$ by finding a particular solution x_0 of $Ax = b$, and defining a new variable $y = x - x_0$. The resulting problem is:

$$\begin{aligned} \text{minimize} \quad & f(y + x_0) \\ \text{subject to} \quad & h(y + x_0) \leq 0, \qquad\qquad\qquad Ax = b \end{aligned}$$

  This problem is equivalent to the original problem, since any solution y of the transformed problem can be converted to a solution x of the original problem by setting $x = y + x_0$.

- **Example 2:** Consider the following convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x_1 + x_2 = 1, \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

  We can eliminate the equality constraint $x_1 + x_2 = 1$ by defining a new variable $y = x_1 + x_2$. The resulting problem is:

$$\begin{aligned} \text{minimize} \quad & f(\frac{y}{2}, \frac{y}{2}) \\ \text{subject to} \quad & y \geq 0, \end{aligned}$$

  This problem is equivalent to the original problem, since any solution y of the transformed problem can be converted to a solution x of the original problem by setting $x_1 = \frac{y}{2}$ and $x_2 = \frac{y}{2}$.

11.2 Results:

The main result on eliminating equality constraints in equivalent convex problems is that the process preserves convexity. This means that if the original problem is convex, then the transformed problem will also be convex. This is important because it allows us to use standard convex optimization methods to solve the transformed problem.

Another important result is that the process of eliminating equality constraints is reversible. This means that we can always reconstruct the solution of the original problem from the solution of the transformed problem. This is useful because it allows us to use the transformed problem to solve the original problem, even if the original problem is difficult to solve directly.

12 Linear Optimization Problems

When the objective and constraint functions are all affine, the problem is called **linear program**.

Linear Program (General Form):

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \leq h, \\ & Ax = b, \quad i = 1, \ldots, p \end{aligned}$$

where $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$

- common to omit d in objective function
- can maximize an affine objective

  $$c^T x + d$$

  by minimizing

  $$-c^T x - d$$

The geometric interpretation of an LP is illustrated in figure below. The feasible set of the LP is a polyhedron $\mathcal{P}$; the problem is to minimize the affine function $c^T x + d$ (or, equivalently, the linear function $c^T x$) over $\mathcal{P}$.

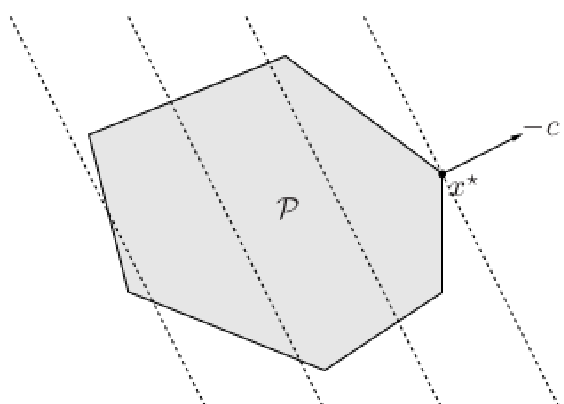


Figure 2: Geometric interpretation of an LP. The feasible set $\mathcal{P}$, which is a polyhedron, is shaded. The objective $c^T x$ is linear, so its level curves are hyperplanes orthogonal to c (shown as dashed lines). The point x^* is optimal; it is the point in $\mathcal{P}$ as far as possible in the direction $-c$.

12.1 Examples:

- **Production planning and scheduling.** Linear optimization can be used to optimize production schedules, minimize costs, and maximize profits.
- **Transportation and logistics.** Linear optimization can be used to optimize routing and scheduling of vehicles, minimize transportation costs, and maximize customer satisfaction.
- **Finance and investment.** Linear optimization can be used to construct optimal portfolios, allocate assets, and manage risks.
- Energy and environment. Linear optimization can be used to optimize energy consumption, reduce emissions, and protect the environment.

12.2 Results:

- **Linear optimization problems are convex problems.** This means that the feasible region of a linear optimization problem is a convex set, and the objective function is a convex function. This property is useful because it allows us to use standard convex optimization methods to solve linear optimization problems.
- Linear optimization problems are solvable in polynomial time. This is in contrast to some other types of optimization problems, which are NP-hard and can be very difficult to solve. The existence of polynomial-time algorithms for solving linear optimization problems is a major reason why linear optimization is such a widely used technique.
- **The duality theorem of linear programming.** The duality theorem of linear programming states that every linear optimization problem has a dual problem, and the optimal value of the primal problem is equal to the optimal value of the dual problem. This theorem is useful for a variety of reasons, including:
  - It can be used to develop new algorithms for solving linear optimization problems.
  - It can be used to prove optimality conditions for linear optimization problems.
  - It can be used to analyze the sensitivity of linear optimization problems to changes in the problem data.

12.3 Illustration:

Diet Problem
In the diet model, a list of available foods is given together with the nutrient content and the cost per unit weight of each food. A certain amount of each nutrient is required per day. For example, here is the data corresponding to a civilization with just two types of grains (G1 and G2) and three types of nutrients (starch, proteins, vitamins): The requirement per day of starch, proteins and

	Starch	Proteins	Vitamins	Cost (dollar/kg)
G1	5	4	2	0.6
G2	7	2	1	0.35

vitamins is 8, 15 and 3 respectively. The problem is to find how much of each food to consume per day so as to get the required amount per day of each nutrient at minimal cost.
Solution:

- Decision variables to use:
  x_1 : number of units of grain G1 to be consumed per day
  x_2 : number of units of grain G2 to be consumed per day

- Write down the objective function:

$$z = 0.6x_1 + 0.35x_2$$

- Constraints:

$$x_1 \geq 0, \quad x2 \geq 0,$$

  Only nonnegative foods can be eaten!

- starch ≥ 8, proteins ≥ 15, vitamins ≥ 3

Optimization Model for Diet Problem

$$\begin{aligned} \text{minimize} \quad & z = 0.6x_1 + 0.35x_2 \\ \text{subject to} \quad & 5x_1 + 7x_2 \geq 8, \\ & 4x_1 + 2x_2 \geq 15 \\ & 2x_1 + x_2 \geq 3 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

13 Quadratic and Quadratically Optimization Problems

Quadratic Program:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^T Px + q^T x + r \\ \text{subject to} \quad & Gx \leq h, \\ & Ax = b, \quad i = 1, \ldots, p \end{aligned} \tag{4}$$

where $P \in S_+^n, G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$

Quadratically Constrained Quadratic Program:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T P_0 x + q^T x + r \\ \text{subject to} & \frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \quad i = 1, \ldots, p \end{array} \tag{5}$$

where $P_i \in S_+^n, \quad i = 1, \ldots, m$, the problem is called a **quadratically constrained quadratic program (QCQP)**.

In a QCQP, we minimize a convex quadratic function over a feasible region that is the intersection of ellipsoids (when $P_i > 0$). Quadratic programs include linear programs as a special case, by taking $P = 0$ in (4). Quadratically constrained quadratic programs include quadratic programs (and therefore also linear programs) as a special case, by taking $P_i = 0$ in (5), for $i = 1, \ldots, m$.

13.1 Examples:

- **Portfolio optimization:** In portfolio optimization, we want to allocate our assets in a way that maximizes our expected return while minimizing our risk. This can be formulated as a QCQP, where the objective function is to maximize the expected return and the constraints are to ensure that the risk does not exceed a certain threshold.
- **Least squares estimation:** In least squares estimation, we want to find the line or curve that best fits a set of data points. This can be formulated as a QCQP, where the objective function is to minimize the sum of the squared residuals and the constraints are to ensure that the line or curve passes through a certain point or passes through a certain set of points.
- **Control system design:** In control system design, we want to design a controller that will keep a system at a desired setpoint. This can be formulated as a QCQP, where the objective function is to minimize the tracking error and the constraints are to ensure that the controller is stable and that the system does not saturate.

13.2 Results:

- QCQPs are NP-hard problems, meaning that there is no known polynomial-time algorithm for solving them exactly. However, there are a number of approximate algorithms that can be used to solve QCQPs.
- One popular approximate algorithm is the interior point method. Interior point methods work by iteratively moving closer to the boundary of the feasible region. They are typically very efficient for solving QCQPs, but they can be sensitive to the initial starting point.
- Another popular approximate algorithm is the sequential quadratic programming (SQP) method. SQP methods work by approximating the QCQP with a sequence of quadratic programs. They are typically less efficient than interior point methods, but they are more robust to the initial starting point.

13.3 Illustration:

- A simple example of a quadratic problem is the following:

$$\text{minimize} \quad f(x) = x^2 + y^2$$

This problem asks us to find the smallest value of the function $f(x) = x^2 + y^2$. The feasible region for this problem is all points in the real plane.

The optimal solution to this problem is $(0, 0)$, which is the minimum point of the parabola.

- Here is a simple example of a QCQP:

$$\begin{aligned} \text{minimize} \quad & f(x) = x^2 + y^2 \\ \text{subject to} \quad & x^2 - y^2 \leq 0 \end{aligned}$$

This problem is the same as the previous example, but with an additional constraint. The constraint states that $x^2 - y^2$ must be less than or equal to zero. This constraint defines a half-plane in the real plane.

The feasible region for this problem is the intersection of the parabola and the half-plane. The optimal solution to this problem is still $(0, 0)$, but it is now more difficult to find because of the additional constraint.

14 Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form:

$$\|Ax + b\|_2 \leq c^T x + d$$

a **second order cone constraint**.

- When $c_i = 0$ for $i = 1, \ldots, m$ then SOCP is equivalent to QCQP.
- If $A_i = 0$ for $i = 1, \ldots, m$, then the SOCP reduces to a LP.

14.1 Examples:

- **Portfolio optimization:** SOCPs can be used to solve the problem of allocating assets in a portfolio to maximize expected return while minimizing risk.
- **Least squares estimation:** SOCPs can be used to solve the problem of finding the best fit line or curve to a set of data points.
- **Control system design:** SOCPs can be used to design controllers that will keep a system at a desired setpoint.
- **Machine learning:** SOCPs can be used to solve a variety of machine learning problems, such as support vector machines and linear regression.

14.2 Results:

SOCPs are NP-hard problems, meaning that there is no known polynomial-time algorithm for solving them exactly. However, there are a number of approximate algorithms that can be used to solve SOCPs.

One popular approximate algorithm for solving SOCPs is the interior point method. Interior point methods work by iteratively moving closer to the boundary of the feasible region. They are typically very efficient for solving SOCPs, but they can be sensitive to the initial starting point.

Another popular approximate algorithm for solving SOCPs is the sequential quadratic programming (SQP) method. SQP methods work by approximating the SOCP with a sequence of quadratic programs. They are typically less efficient than interior point methods, but they are more robust to the initial starting point.

14.3 Illustration:

Consider a transportation optimization problem where you need to minimize transportation costs subject to certain constraints. You have n cities, and you want to find the optimal amount of goods to transport between them. Your objective is to minimize transportation costs.

Objective Function:
Minimize the total transportation cost, which is a linear function of the amount of goods to be transported between cities, represented by the vector x.

Constraints:
1. Each city must receive a certain minimum amount of goods, represented by second-order cone constraints: $\|A_i x + b_i\| \leq c_i^T x + d_i, \quad i = 1, \ldots, m$

2. Each of these constraints ensures that the goods transported to city i satisfy certain minimum requirements.

3. The total amount of goods transported from all cities must be equal to a certain demand, represented as an equality constraint: $Fx = g$

This constraint ensures that the total demand for goods is met.

3. Non-negativity constraints on the amount of goods to be transported: $x \geq 0$ This is an illustrative example of an SOCP problem in the context of transportation optimization. It involves linear objective and constraint functions but includes second-order cone constraints to model the minimum requirements of goods to be transported to each city. The goal is to find the optimal allocation of goods that minimizes transportation costs while meeting the city's requirements.

15 Robust Linear Programming

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m \end{array}$$

where there are uncertainty in c, a_i , b_i.

- To simplify assume c and b_i are fixed.
- Assume a_i lies in the given ellipsoids:

$$a_i \in \mathcal{E}_i = \{\overline{a}_i + P_i u \quad | \| \quad \mathbf{u}\|_2 \leq 1\}$$

$P_i \in \mathbb{R}^{n \times n}$

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \text{for all} \quad a_i \in \mathcal{E} \quad i = 1, \ldots, m \end{array}$$

15.1 Examples:

- **Portfolio optimization:** The decision variables a_i could represent the amount of money invested in different assets, and the constraints $ar < b$ could represent the risk constraints. The uncertainty in the problem could represent uncertainty in the future returns of the assets.
- **Production planning:** The decision variables a_i could represent the amount of different products to produce, and the constraints $ar < b$ could represent the capacity constraints. The uncertainty in the problem could represent uncertainty in demand.

15.2 Results:

There are a number of different ways to solve robust linear programs. One common approach is to use a scenario-based approach. In this approach, the problem is reformulated as a deterministic linear program with a large number of scenarios. Each scenario represents a different possible realization of the uncertainty, and the objective is to find a solution that is feasible for all of the scenarios.

Another approach to solving robust linear programs is to use a chance-constrained approach. In this approach, the constraints are modified to ensure that the probability of violating them is less than a certain threshold. The objective is then to minimize the expected value of the objective function.

15.3 Illustration:

Consider a company that wants to optimize its supply chain decisions for a particular product. The goal is to determine how much of the product to produce at different factories and how much to ship to various distribution centers.

Objective: Minimize the total cost, including production costs and transportation costs.

Decision Variables:
- x_i represents the quantity of the product to be produced at factory i
- y_{ij} represents the quantity of the product to be shipped from factory i to distribution center j.

Uncertainty in Demand and Production Costs: The company faces uncertainty in two key parameters: demand and production costs. Demand at each distribution center and production costs at each factory are uncertain. The uncertainty in demand and production costs is represented as follows:
1. **Demand Uncertainty:**

The demand at each distribution center j is uncertain and can vary within a range: $d_j \pm \Delta d_j$, where d_j is the nominal demand, and Δd_j represents the uncertainty.

2. **Production Cost Uncertainty:**

The production cost at each factory i is uncertain and can vary within a range: $c_i \pm \Delta c_i$, where c_i is the nominal cost, and Δc_i represents the cost uncertainty.

Robust Optimization Model: The company wants to make supply chain decisions that are robust to these uncertainties. The robust linear programming model could be formulated as follows:

Objective: Minimize the expected cost, considering the worst-case demand and cost scenarios within the uncertainty sets.

Constraints:

- Production Capacity Constraints: For each factory i, ensure that the produced quantity x_i does not exceed the maximum production capacity.
- Shipment Constraints: Ensure that the quantity shipped from factory i to distribution center j does not exceed the production capacity and meets the demand.
- Uncertainty Set Constraints: For each distribution center j, the total demand from all factories should be at least $d_j - \Delta d_j$
- Cost Uncertainty Constraints: For each factory i, ensure that the total production cost does not exceed $c_i + \Delta c_i$

Non-Negativity Constraints: Ensure that all decision variables are non-negative.

Objective Function: Minimize the expected total cost, which is a linear combination of production costs and transportation costs.

In this example, robust linear programming helps the company make supply chain decisions that account for uncertainties in both demand and production costs. By considering the worst-case scenarios within the defined uncertainty sets, the company aims to ensure that their supply chain plan remains effective even in the face of unexpected changes in demand and costs.

16 Semidefinite Programming (SDP)

Semidefinite Programming:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + x_2 F_2 + \ldots + x_n F_n \leq 0 \\ & Ax = b, \end{aligned}$$

where $G, F_1, \ldots, F_n \in S^k$, and $A \in \mathbb{R}^{p \times n}$.

If the matrices $G, F_1, ..., F_n$ are diagonals then LMI (Linear Matrix Inequality) reduces to a set of n linear inequalities, and SDP becomes LP.

16.1 Examples:

- Finding the maximum cut of a graph
- Truss topology optimization
- Phase retrieval
- Sensor network localization
- Machine learning problems such as support vector machines and max-cut problems

16.2 Results:

- SDP is a powerful tool for solving a wide variety of problems in optimization, control, and machine learning.
- SDP problems can be solved efficiently using interior-point methods.
- SDP problems have been shown to be NP-hard, but there are many approximation algorithms that can be used to find good solutions.

16.3 Illustration:

One classic and illustrative example of an SDP problem is the Max-Cut problem, which is commonly encountered in graph theory and network analysis.

The Max-Cut problem is defined as follows:

Problem Statement: Given an undirected graph $G = (V, E)$, where V is the set of vertices and E is the set of edges, find a partition of the vertices into two sets, V_1 and V_2, such that the number of edges crossing the partition (i.e., with one endpoint in V_1 and the other in V_2) is maximized.

To formulate the Max-Cut problem as an SDP, you can define a decision variable X, which is a symmetric matrix of size $|V|x|V|$, where $|V|$ is the number of vertices. The SDP objective is then to maximize the following expression: Maximize: $\mathrm{Tr}(C \cdot X)$

Subject to:

1. X is a positive semi-definite matrix (X is a symmetric matrix, and all its eigenvalues are non-negative).
2. X is a diagonal matrix with 1's on the diagonal for vertices in V_1 and -1's on the diagonal for vertices in V_2 (to represent the partition).

Here, C is the adjacency matrix of the graph G. The elements of C represent the weights of the edges in the graph. $\mathrm{Tr}(C \cdot X)$ represents the total weight of edges crossing the partition.

The SDP formulation for the Max-Cut problem can be solved using SDP solvers like CVX, and it provides an optimal partition of the graph's vertices into two sets, such that the number of cut edges is maximized.

17 Duality

17.1 The Lagrange dual function

17.1.1 The Lagrangian

We consider an optimization problem in the standard form:

Optimization Problem(Standard form):

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

with variable $x \in \mathbb{R}^n$. We assume its domain D $= \bigcap_{i=0}^{m} \mathrm{dom} f \quad \cap \quad \bigcap_{i=1}^{p} \mathrm{dom} h$ is nonempty, and denote the optimal value f_0 by p^* . We do not assume the problem is convex.

The basic idea in Lagrangian duality is to take the constraints into account by augmenting the objective function with a weighted sum of the constraint functions. We define the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with the standard optimization problem as:

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

with $domL = D \times \mathbb{R}^m \times \mathbb{R}^p$.

- We refer to λ_i as the **Lagrange multiplier** associated with the i^{th} inequality constraint $f_i(x) \leq 0$;
- Similarly we refer to v_i as the Lagrange multiplier associated with the i_{th} equality constraint $h_i(x) = 0$.
- The vectors λ and v are called the **dual variables** or **Lagrange multiplier vectors** associated with the standard optmization problem.

17.1.2 The Lagrange dual function

We define the **Lagrange dual function (or just dual function)** $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x : for$\lambda \in \mathbb{R}^m, v \in \mathbb{R}^p$,

$$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} (f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x))$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, v), it is concave, even when the standard problem is not convex.

17.2 Examples:

Here are some examples of how the Lagrangian and the Lagrange dual function can be used to solve constrained optimization problems:

- **Linear programming:** Linear programming problems can be solved using the Lagrangian and the Lagrange dual function. The Lagrange dual function for a linear programming problem is a concave function, which makes it easy to solve.
- **Quadratic programming:** Quadratic programming problems can also be solved using the Lagrangian and the Lagrange dual function. The Lagrange dual function for a quadratic programming problem is a quadratic function, which is also easy to solve.
- **Nonlinear programming:** Nonlinear programming problems can be solved using the Lagrangian and the Lagrange dual function, but the dual problem is often more difficult to solve than the primal problem.

17.3 Results:

Here are some important results about the Lagrangian and the Lagrange dual function:

- The strong duality theorem states that the optimal value of the primal problem is equal to the optimal value of the dual problem, under certain conditions.
- The weak duality theorem states that the optimal value of the dual problem is always less than or equal to the optimal value of the primal problem.
- The Karush-Kuhn-Tucker (KKT) conditions are a set of necessary and sufficient conditions for optimality in a constrained optimization problem. The KKT conditions can be expressed in terms of the Lagrangian and the Lagrange multipliers.

17.4 Illustration:

Imagine you are a shipping company and you want to minimize the cost of shipping goods from point A to point B. You have a fleet of trucks and vans, each of which has a different capacity and cost to operate. You also have a constraint that you must deliver all of the goods within a certain amount of time.

This can be formulated as a constrained optimization problem as follows:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} c_i x_i \\
\text{subject to} \quad & \sum_{i=1}^{n} w_i x_i \geq G \\
& \sum_{i=1}^{n} t_i x_i \leq T
\end{aligned}
$$

where:

- n is the number of trucks and vans in your fleet
- c_i is the cost to operate truck or van i
- x_i is the number of goods to be shipped by truck or van i
- w_i is the capacity of truck or van i
- G is the total weight of the goods to be shipped
- t_i is the time it takes truck or van i to travel from point A to point B
- T is the maximum amount of time allowed for delivery

The Lagrangian for this problem is defined as follows:

$$L(x, \lambda, \mu) = \sum_{i=1}^{n} c_i x_i + \lambda(\sum_{i=1}^{n} w_i x_i - G) + \mu(\sum_{i=1}^{n} t_i x_i - T)$$

where:

- λ is the Lagrange multiplier for the capacity constraint
- μ is the Lagrange multiplier for the time constraint

The Lagrange dual function is defined as follows:

$$g(\lambda, \mu) = inf_x L(x, \lambda, \mu)$$

The Lagrange dual function is a concave function of λ and μ, so it is relatively easy to solve the dual problem.

Once the dual problem is solved, the primal variables x_i can be recovered using the following KKT conditions:

$$\sum_{i=1}^{n} w_i x_i = G$$
$$\sum_{i=1}^{n} t_i x_i = T$$

This illustration shows how the Lagrangian and the Lagrange dual function can be used to solve a complex constrained optimization problem.

18 Link to the scribe:

Click here to edit this scribe