Optimization for Machine Learning (CSL4010)

Dr. Md Abu Talhamainuddin Ansary
Department of Mathematics, IIT Jodhpur

PPT 3

Linear and Nonlinear Optimization Problems

Linear programming problem

- If f, g_i for all i and h_j for all j are affine, then (P) is said to be a linear programming (LP).
- Generalized form of an LP:

$$\begin{aligned} \min c^T x & \\ s.\,t.\; Ax \leq & \quad b \\ A_{eq}x = & \quad b_{eq} \\ x \geq 0 & \end{aligned}$$

- If the LP has no equality constraints, it is called an inequality form LP.
- If there exists any x_i in an LP which is unrestricted in sign, then x_i is expressed as $x_i = x_i' - x_i''$ where $x_i', x_i'' \geq 0$.

Linear programming problem

- If f, g_i for all i and h_j for all j are affine, then (P) is said to be a linear programming (LP).
- Generalized form of an LP:

$$\begin{aligned} \min c^T x & \\ s.\ t.\ Ax \leq & \quad b \\ A_{eq}x = & \quad b_{eq} \\ x \geq 0 & \end{aligned}$$

- If the LP has no equality constraints, it is called an inequality form LP.
- If there exists any x_i in an LP which is unrestricted in sign, then x_i is expressed as $x_i = x_i' - x_i''$ where $x_i', x_i'' \geq 0$.
- Leonid Vitalyevich Kantorovich is known as founder of LPP who won Nobel prize 1975 in economics for his contribution in optimal allocation of scarce resources.

Diet problem

- Suppose a person takes n different different foods. Each food contains following amount different nutrients.

nutrient	$food_1$	$food_2$	...	$food_n$
carbohydrate	a_{11}	a_{12}	...	a_{1n}
protein	a_{21}	a_{22}	...	a_{2n}
fat	a_{31}	a_{32}	...	a_{3n}
vitamin	a_{41}	a_{42}	...	a_{4n}

- Suppose a person requires at least b_1, b_2, b_3, and b_4 amount of different nutrients.
- Suppose price of each food item is c_i.

- If a person want to take x_i amount of i^{th} food with the cheapest diet that contains all nutritional requirements. Then we have to solve the following LP:

$$\begin{aligned} \min \quad & c_1x_1 + c_2x_2 + \cdots + c_nx_n & & \\ s.\ t.\ & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & \geq \quad & b_1 \\ & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & \geq \quad & b_2 \\ & a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n & \geq \quad & b_3 \\ & a_{41}x_1 + a_{42}x_2 + \cdots + a_{4n}x_n & \geq \quad & b_4 \\ & x_1, x_2, \ldots, x_n & \geq \quad & 0 \end{aligned}$$

- The above problem can be written in matrix form as

$$\begin{aligned} \min c^T x & \\ Ax \geq & \quad b \\ x \geq 0 & \end{aligned}$$

Chebyshev centre of a polyhedron

- Consider the polyhedron $\mathcal{P} + \{x | {a^i}^T x \leq b_i, \; i = 1, 2, \dots, m\}$.
- Find the largest Euclidean ball that lie in $\mathcal{P}$.
- Represent the ball $B = \{x_c + u | \|u\|_2 \leq r\}$. Here x_c and r are unknown.
- B must lie in half-space ${a^i}^T x \leq b_i$ for all i. i.e.

$$ {a^i}^T (x_c + u) \leq b_i. $$

- Since $\sup\{{a^i}^T u | \|u\| \leq r\} = r\|a\|_2$, the constraints can be written as

$$ {a^i}^T x_c + r\|a^i\|_2 \leq b_i $$

  for $i = 1, 2 \dots, m$.
- Hence the problem become

$$ \begin{array}{rrcl} & \max \; r & & \\ s.\,t. & {a^i}^T x_c + r\|a^i\|_2 & \leq & b_i \quad \forall\, i = 1, 2, \dots, m. \end{array} $$

Chebyshev inequalities

- Suppose $X = \{x^1, x^2, \ldots, x^n\}$ be a set of discrete random variable with $P(x = x_i) = p_i$. Then $0 \le p_i \le 1$, $\sum_{i=1}^n p_i = 1$.
- Expectation of $f(x)$ is $E(f(x)) = \sum_{i=1}^n p_i f(x^i) = f^T p$, where $f = (f(x^i), i = 1, 2, \ldots, n)^T$ and $p = (p_1, p_2, \ldots, p_n)^T$.
- Suppose p is not known to us, but we know bounds of $E(f_j(x))$ for $j = 1, 2, \ldots, m$.
- we have to find $\min E(f_0(x))$,.
- The problem is

$$\begin{aligned} \min \ & f_0^T p \\ s.t. l_i \le f_i^T p & \le u_i \ \ i = 1, 2, \ldots, m \\ \bar{1}^T p & = 1 \\ p \ge 0 & \end{aligned}$$

Transportation problem

- Consider the problem

Destination

Source		D1	D2	D3	D4	Supply(s_i)
	O1	C_{11}	C_{12}	C_{13}	C_{14}	s_1
	O2	C_{21}	C_{22}	C_{23}	C_{24}	s_2
	O3	C_{31}	C_{32}	C_{33}	C_{34}	s_3
	O4	C_{41}	C_{42}	C_{43}	C_{44}	s_4
Demand (d_j):		d_1	d_2	d_3	d_4	

Figure 1: Transportation problem

- Objective function is min $\sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij}x_{ij}$
- Constraints for sources $\sum_{j=1}^{4} x_{ij} \leq s_i$, $i = 1, 2, 3, 4$.
- Constraints for destination $\sum_{i=1}^{4} x_{ij} \leq d_j$, $j = 1, 2, 3, 4$.

Network flow problem

- Consider the network flow problem. In this problem electricity is transferred from nod 1 (source) to node 5 (sink). There are some intermediate nodes 2,3,4.

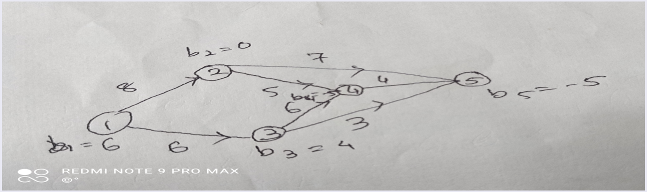

Figure 2: Network flow problem

- Some nodes are connected and each connected node has cost to transfer electricity.
- Each node has some electricity stored (b_i). $b_i > 0$ implies i th node has surplus and $b_i < 0$ implies this node has some deficit.
- We have to find minimum cost for supplying power from node 1 to node 5.

Network flow problem contd...

- Suppose x_{ij} be the amount of electricity should pass from node i to node j.
- Hence the objective function is $f(x) = 8x_{12} + 6x_{13} + 5x_{24} + 7x_{25} + 6x_{34} + 3x_{35} + 4x_{45}$
- Since node 1 has 6 unit electricity so $x_{12} + x_{13} = 6$.
- For node 2, $x_{24} + x_{25} = x_{12}$ i.e. $-x_{12} + x_{24} + x_{25} = 0$.
- Similarly we can construct constraints for other nodes.

Network flow problem contd...

- So the optimization problem become

$$\begin{aligned} \min \quad & 8x_{12} + 6x_{13} + 5x_{24} + 7x_{25} + 6x_{34} + 3x_{35} + 4x_{45} \\ s.t. \quad & x_{12} + x_{13} = 6 \\ & -x_{12} + x_{24} + x_{25} = 0 \\ & -x_{13} + x_{34} + x_{35} = 4 \\ & -x_{24} - x_{34} + x_{45} = -5 \\ & -x_{25} - x_{35} - x_{45} = -5 \\ & x_{ij} \geq 0 \end{aligned}$$

- In this problem number of variables is equal to the number of edges and number of constraints is equal to the number of nodes.
- One can observe that coefficients of x_{ij} in i^th constraint is 1 then coefficient of x_{ij} in j^{th} constraint is -1.
- In this problem we assume $\sum b_i = 0$.
- In some cases each x_{ij} may be given a lower bound l_{ij} and upper bound u_{ij}. In this case box constraints are included in LP.

Shortest path problem

- Consider paths form source 1 to destination 7. There are some intermediate nodes. We have to find shortest path from 1 to 7.

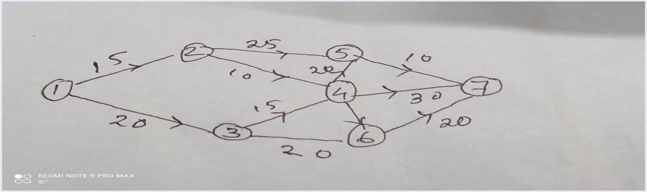

Figure 3: Shortest path problem

- While choosing shortest path either we choose one path or not. So decision variables x_{ij} values will be

$$x_{ij} = \begin{cases} 1 \text{ if } ij \text{ path is selected} \\ 0\text{otherwise} \end{cases}$$

Shortest path problem contd...

- Hence the objective function is $f(x) = 15x_{12} + 20x_{13} + 10x_{24} + 25x_{25} + 14x_{34} + 20x_{36} + 20x_{45} + 15x_{46} + 30x_{47} + 10x_{57} + 20x_{67}$
- Since we starts from node we, we have to either choose x_{12} or x_{13}. So first constraint is $x_{12} + x_{13} = 1$
- For node 2, $x_{24} + x_{25} = x_{12}$ i.e. $-x_{12} + x_{24} + x_{25} = 0$.
- Similarly we can construct constraints for other intermediate nodes.
- Since we must reach node 7 (destination)so the constraint for node 7 is $x_{47} + x_{57} + x_{67} = 1$ i.e $-x_{47} - x_{57} - x_{67} = -1$

Shortest path problem contd...

- So the optimization problem become

$$
\begin{aligned}
\min \quad 15x_{12} + 20x_{13} + 10x_{24} + 25x_{25} \quad & + \\
14x_{34} + 20x_{36} + 20x_{45} \quad & + \quad 15x_{46} + 30x_{47} + 10x_{57} + 20x_{67} \\
s.t. \qquad x_{12} + x_{13} \quad & = \quad 1 \\
-x_{12} + x_{24} + x_{25} \quad & = \quad 0 \\
-x_{13} + x_{34} + x_{36} \quad & = \quad 0 \\
-x_{24} - x_{34} + x_{45} + x_{46} + x_{47} \quad & = \quad 0 \\
-x_{25} - x_{45} + x_{47} \quad & = \quad 0 \\
-x_{36} - x46 + x_{67} \quad & = \quad 0 \\
-x47 - x_{57} - x_{67} \quad & = \quad -1 \\
x_{ij} \in \{0, 1\} \quad &
\end{aligned}
$$

- In this problem number of variables is equal to the number of edges and number of constraints is equal to the number of nodes.

Shortest path problem contd...

- **Definition:** An square matrix ($A_{n \times n}$) with integer entries is said to be unimodular if $det(A) = \pm 1$.
- A matrix $A \in \mathbb{Z}^{m \times n}$ is said to be totally unimodular if every square submatrix of A is unimodular or is singular (i.e. has determinant 0). (A $k \times k$ submatrix of A is obtained by removing any $m - k$ rows and $n - k$ columns of A .)
- Let $A \in \{-1, 0, 1\}^{m \times n}$. If every column of A has at most one 1 and at most one -1 , then A is totally unimodular.
- Consider the LP

$$\begin{aligned} \min \quad & c^T x \\ s.\ t.\ Ax &= b, x \geq 0 \end{aligned}$$

- If A is an unimodular matrix then solution of the above LP (if exists) is an integer solution.
- One can observe that coefficient matrix of shortest path problem is a totally unimodular matrix. Hence in state of solving $x_{ij} \in \{0, 1\}$, we can solve the LP to find optimal solution.

Assignment problem

- There are n persons and n tasks. Each person is to be assigned to exactly one task such that no two persons are assigned to the same task. The cost of person i doing task j is given by c_{ij}.
- The problem is to find an assignment of persons to tasks that minimizes total cost.
- Since each person is assigned only one task $\sum_{j=1}^{n} x_{ij} = 1$ for all i.
- Since one task is assigned to only one person task $\sum_{i=1}^{n} x_{ij} = 1$ for all j.
- Hence the assignment problem become

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1, \quad i = 1, 2, \ldots, n \\ \sum_{i=1}^{n} x_{ij} &= 1 \quad j = 1, 2, \ldots, n \end{aligned}$$

$$x_{ij} \in \{0, 1\}$$

Assignment problem contd...

- If we write last n constraints as $\sum_{i=1}^{n} -x_{ij} = -1$.
- Then coefficient matrix of the above problem becomes a unimodular matrix.
- Hence the assignment problem can be solved by solving the following LP:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1, \quad i = 1, 2, \ldots, n \\ \sum_{i=1}^{n} -x_{ij} &= -1 \quad j = 1, 2, \ldots, n \\ x_{ij} &\geq 0 \end{aligned}$$

Quadratic programming problem

- If the objective function in (P) is quadratic and constraints are affine then it is said to be a quadratic problem (QP).
- Generalized form of a QP:

$$\begin{aligned} \min \frac{1}{2}x^T Qx + c^T x & \\ s.\,t.\, Ax \leq & \quad b \\ A_{eq}x = & \quad b_{eq} \end{aligned}$$

- Clearly f will be a convex function iff Q is positive semi-definite.

Constrained least square problems

- A least square problem of form $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$ is an unconstrained problem.
- If linear constraints are added then it is called constrained least square problem.
- Consider the least square problems with lower and upper bounds on the variables. i.e.

$$\begin{aligned} \min \frac{1}{2}\|Ax - b\|^2 &= \frac{1}{2}x^T A^T Ax - A^T bx + \frac{1}{2}b^T b \\ s.t. l_i \leq x_i \leq b_i & \end{aligned}$$

- The above problem is a QP.
- An optimization problem of the form

$$\begin{aligned} \min f(x) \\ s.t. l_i \leq x_i \leq b_i \end{aligned}$$

is said to be a box constrained problem.

Portfolio optimization

- The Markowitz covariance model uses t minimize the risk of a portfolio, represented by its variance for at least certain amount of expected return.
- This model can be represented as

$$\begin{aligned} \min \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} K_{ij} x_i x_j & \\ \text{s. t. } \Sigma_{i=1}^{n} \mu_i x_i & \geq r \\ \Sigma_{i=1}^{n} x_i = 1 & \\ x_i \geq 0, \ \ i \in \Lambda_n & \end{aligned}$$

where $K_{n \times n}$ is the covariance matrix.

- The Standard or unit second-Order (convex) cone of dimension $n+1$ is defined as

$$C_{n+1} = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^{n+1}, x \in \mathbb{R}^n, t \in \mathbb{R} | \|x\|_2 \leq t \right\}$$

- An inequality constraint of the form $\|Ax+b\| \leq c^T x + d$ is said to be a g a second-Order cone constraint.
- The set of points satisfying a second-Order cone constraint is the inverse image of the unit second-Order cone under an affine mapping. i.e.

$$\|Ax+b\|_2 \leq c^T x + d \Leftrightarrow \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix} \in C_{n+1}$$

- An optimization problems with second order cone constraints is said to be second order cone programming. General form of a second order cone constrained optimization problem is

$$\begin{aligned} (P_c) : \min \quad & f^T x \\ s.t. \quad \|A_i x + b_i\|_2 & \leq c_i^T x + d_i \quad i = 1, 2, \ldots, m \\ A_{eq} x & = b_{eq} \end{aligned}$$

- If in P_c $A_i = 0$ for all i then P_c reduced to a linear programming problem

- **Quadratically constrained quadratic problem(QCQP):** (P) is said to be a QCQP if objective function as well as all constraint functions are convex quadratic function. i.e. A QCQP is of the form

$$\begin{aligned}(P_{QCQP}): \quad \min x^T Qx + c^T x + r_0 & \\ s.t. \frac{1}{2}x^T A_i x + b_i^T x + r_i & \leq 0 \\ A_{eq}x = b_{eq} &\end{aligned}$$

  where Q and each A_i are symmetric positive definite matrix.
- If Q is a diagonalize matrix then Q can be written as $Q = PDP^{-1}$ where $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, λ_i are eigenvalues of Q.
- Since Q is positive definite $\lambda_i > 0$. Then Q can be written as $Q = (PD^{1/2}P^{-1}).(PD^{1/2}P{-}1)$, where $D^{1/2} = diag(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2})$.
- Define $Q^{1/2} = PD^{1/2}P^{-1}$. Then objective function in $P_{QCQP})$ can be written as $\frac{1}{2}\|Q^{1/2}x + Q^{-1/2}c\|^2 + r_0 - 1/2c^T Qc$.

- Similarly constraints can be written as $\frac{1}{2}\|A_i^{1/2}x + A_i^{-1/2}b_i\|^2 + r_i - 1/2 b_i^T A_i^{1/2} b_i \leq 0.$
- Thus (P_{QCQP}) can be written as

$$
\begin{aligned}
(P_{QCQP}): \quad \min t & \\
s.t. \|Q^{1/2}x + Q^{-1/2}c\| & \leq t \\
\|A_i^{1/2}x + A_i^{-1/2}b_i\| & \leq (b_i^T A_i b_i - 2r_i)^{1/2} \\
A_{eq}x & = b_{eq}
\end{aligned}
$$

- **Discrete Optimization:** If feasible set of an optimization problem is a discrete set then it is called a discrete optimization.

$$\begin{aligned} \min \quad f(x) & \\ s.t. g_i(x) &\leq 0 \\ h_j &= 0 \\ x \in I & \end{aligned}$$

  where I is a discrete set of points.

- **Integer programming problem:** If feasible set of an optimization problem is the set of integers then it is called a discrete optimization.

$$\begin{aligned} \min \quad f(x) & \\ s.t. g_i(x) &\leq 0 \\ h_j &= 0 \\ x \in \mathbb{Z}. & \end{aligned}$$

- **Knapsack problem:** If feasible set of an optimization problem is the set $\{0, 1, 2, \ldots, n\}$ then it is called knapsack optimization problem.

$$\begin{aligned} \min \quad f(x) & \\ s.t. g_i(x) & \leq 0 \\ h_j & = 0 \\ x \in \{0, 1, 2, \ldots, n\}. & \end{aligned}$$

- $0 - 1$ **Knapsack problem:** If feasible set of an optimization problem is the set $\{0, 1\}$ then it is called $0 - 1$ knapsack optimization problem.

$$\begin{aligned} \min \quad f(x) & \\ s.t. g_i(x) & \leq 0 \\ h_j & = 0 \\ x \in \{0, 1\}. & \end{aligned}$$

- Consider the LP

$$\min \quad \max\{c_1^T x, c_2^T x, \ldots, c_m^T x\}$$
$$\begin{aligned} s.t. Ax &\leq b \\ A_{eq}x &= b_{Eq} \end{aligned}$$

- Clearly this is a convex optimization problem but objective function is non differentiable.
- Suppose $t = \max\{c_1^T x, c_2^T x, \ldots, c_m^T x\}$. Then the above problem can be written as

$$\min_{t,x} \quad t$$
$$\begin{aligned} c_i^T x &\leq t, \quad i = 1, 2, \ldots, m \\ s.t. Ax &\leq b \\ A_{eq}x &= b_{Eq} \end{aligned}$$

Which is a smooth LP.

- Thus introducing new variable t we can convert the piecewise smooth problem to a smooth problem.

- Similarly consider a nonlinear problem

$$\begin{aligned} \min \quad \max\{f_1(x), f_2(x), \ldots, f_m(x)\} \\ s.t. g_i(x) &\leq 0 \\ a_j^T x &= b_j \end{aligned}$$

  where $f_1, f_2, \ldots, f_m$ are convex functions. Clearly this is a nonsmooth convex optimization problem.
- Introducing a new variable t we can convert this to

$$\begin{aligned} \min_{t,x} \quad & t \\ s.t. \quad f_l(x) &\leq t, \quad l = 1, 2, \ldots, m \\ g_i(x) &\leq 0 \\ a_j^T x &= b_j \end{aligned}$$

- Consider a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Define $dom\ f = \{x \in \mathbb{R}^n | f(x) < \infty\}$.
- $\xi \in \mathbb{R}^n$ is said to be a subgradient of f at $x \in dom\ (f)$ iff

$$f(y) - f(x) \geq \xi^T(y - x) \tag{1}$$

  for all $y \in \mathbb{R}^n$.
- Since $f(y) = \infty$ for all $y \notin dom\ f$, inequality (1) holds for all $y \notin dom\ (f)$.
- Hence we can say $\xi \in \mathbb{R}^n$ is said to be a subgradient of f at $x^* \in dom\ (f)$ iff

$$f(y) - f(x^*) \geq \xi^T(y - x^*)$$

  for all $y \in dom\ (f)$.
- Consider $f(x) = |x|$ and $x^* = 0$. Once can observe that $|y| - 0 \geq \xi(y - 0)$ holds for any $\xi \in [-1, 1]$. Thus any real number in $[-1, 1]$ is a subgradient of f at x.

- Subdifferential of a convex function at $x^* \in dom\, f$ is defined as

$$\partial f(x^*) = \{\xi \in \mathbb{R}^n | f(y) - f(x) \geq \xi^T (y - x) \ \ \forall\, y \in dom\,(f)\} \tag{2}$$

  i.e. collection of all subgradients is called the subdifferential.
- Consider $f(x) = |x|$ and $x^* = 0$. Then $\partial f(0) = [-1, 1]$.
- If f is continuous at x then $\partial f(x)$ is a compact set (closed and bounded).
- If $f(x) = f_1(x) + f_2(x)$ then $\partial f(x^*) = \partial f_1(x^*) + \partial f_2(x^*)$.
- If $f(x) = \alpha f_1(x)$, $\alpha > 0$ then $\partial f(x^*) = \alpha \partial f_1(x^*)$.
- If $f(x) = \max\{f_1(x), f_2(x), \ldots, f_m(x)\}$. Then

$$\partial f(x^*) = Co\{\nabla g_i(x^*) | i \in \{1, 2, \ldots, m\}, f(x^*) = f_i(x^*).$$

- If $x^* \notin dom\,(f)$ then we define $\partial f(x^*) = \emptyset$.
- If f is differentiable at x^* then $\partial f(x^*) = \{\nabla f(x^*)\}$.
- Suppose $0 \in \partial f(x^*)$ for some $x^* \in dom(f)$ then $f(y) - f(x^*) \geq 0^T (y - x^*)$. This implies $f(y) \geq f(x^*)$ for all $y \in dom\,(f)$. i.e. x^* is a global minima of f.
- Thus x^* is a global minima of f iff

$$0 \in \partial f(x^*).$$