0.1

N. L. P.

Katta G. Murty, IOE 611 Lecture slides

Introductory Lecture

NONLINEAR PROGRAMMING (NLP) deals with optimization models with at least one nonlinear function.

NLP does not include *everything other than linear*.

NLP does not include **IP** or **Discrete Opt**. Needs special **Enumerative Techniques**.

NLP, also called Continuous Optimization or Smooth Optimization. Models of following form:

$$
\begin{aligned}
\text{Minimize} \quad & \theta(x) \\
\text{s. to} \quad h_i(x) &= 0, \quad i = 1 \text{ to } m \\
g_p(x) &\geqq 0, \quad p = 1 \text{ to } t
\end{aligned}
$$

All functions $\theta(x), h_i(x), g_p(x)$ assumed Smooth Functions.

Smooth Function = one with all derivatives.

Beyond 2nd derivatives, impractical particularly when many variables. So, for us **smooth** means:

$$\begin{array}{rl} \text{Continuously differentiable} & \text{if using gradients only} \\ \text{Twice continuously differentiable} & \text{if using Hessians} \end{array}$$

Inequality constraints include lower and upper bounds on decision variables: $\ell_j \leqq x_j \leqq u_j$.

$\nabla f(\bar{x})$ denotes $(\frac{\partial f(x)}{\partial x_j})$ at $x = \bar{x}$ written as row vector. Also called the **gradient** of $f(x)$ at $\bar{x}$.

$\nabla^2_{xx} f(\bar{x})$ denotes $(\frac{\partial^2 f(x)}{\partial x_i \partial x_j})$, $n \times n$ **Hessian matrix** of $f(x)$ at $\bar{x}$,

If $g(x) = (g_1(x), \ldots, g_m(x))^T$ is vector of functions, then $\nabla g(x) = (\frac{\partial g_i(x)}{\partial x_j} : i = 1 \text{ to } m,\ j = 1 \text{ to } n)$, the $m \times n$ **Jacobian matrix** of $g(x)$ at $\bar{x}$. Each row vector in $\nabla g(x)$ is the gradient vector of one function in $g(x)$.

QUADRATIC FUNCTIONS: Simplest nonlinear functions

A Quadratic Form in $x = (x_1, \ldots, x_n)^T$ is of form $f(x) = \Sigma_{i=1}^n q_{ii}x_i^{\mathbf{2}} + \Sigma_{i=1}^n \Sigma_{j=i+1}^n q_{ij}x_ix_j$.

Define **Square symmetric matrix** $D = (d_{ij})$ of order n where

$$d_{ii} = q_{ii} \quad \text{for } i = 1 \text{ to } n$$

$$d_{ij} = d_{ji} = \tfrac{1}{2}q_{ij} \quad \text{for } i \neq j, j > i$$

Then $f(x) = x^T D x$

Example: $h(x) = 81x_1^{\mathbf{2}} - 7x_2^{\mathbf{2}} + 5x_1x_2 - 6x_1x_3 + 18x_2x_3$.

A Quadratic Function is of form $Q(x) = x^T D x + cx + c_0$.

A square matrix M of order n, whether symmetric or not, is:

Positive semidefinite (PSD) if $x^T M x \geqq 0$ for all $x \in R^n$

Positive definite (PD) if $x^T M x > 0$ for all $x \neq 0$

Convex, Concave Functions:

A function $f(x)$ defined over R^n, or some convex subset of R^n, is Convex Function iff for all x^1, x^2 in that set, and all

$0 \leq \alpha \leq 1$,

$$f(\alpha x^1 + (1-\alpha)x^2) \leq \alpha f(x^1) + (1-\alpha) f(x^2)$$

Inequality called Jensen's Inequality after Danish mathematician who defined it in 1905.

Geometric interpretation as *function lying beneath every chord*

Function $f(x)$ Concave if above inequality holds other way, i.e., $-f(x)$ is convex.

Some Properties of Convex Functions

1. Nonnegative Combinations of convex functions convex.

2. If $f(x)$ is convex defined on convex set Γ, then for all $x^1, \ldots, x^r \in \Gamma$ and $\alpha_1, \ldots, \alpha_r \geq 0$ satisfying $\Sigma_{i=1}^r \alpha_i = 1$,

$$f(\sum_{i=1}^{r} \alpha_i x^i) \leq \sum_{i=1}^{r} \alpha_i f(x^i)$$

3. $f(x)$ convex iff its Epigraph is a convex set.

4. If $f(x)$ convex, for all α, the set $\{x : f(x) \leq \alpha\}$ is a convex set. Converse not true.

5. Pointwise supremum function of convex functions convex.

6. Differentiable function defined on real line convex iff its 1st derivative is a monotonic increasing function, i.e., iff its 2nd derivative is nonnegative function.

7. Quadratic function defined over R^n covex over R^n iff its Hessian is PSD.

Quadratic function whose Hessian not PSD, may be convex over a subspace of R^n.

8. Twice continuously differentiable $f(x)$ defined over R^n is convex iff its Hessian PSD for all x.

9. Gradient Support Inequality: Differentiable function $f(x)$ defined over R^n is convex iff for all $\bar{x}$

$$f(x) \geq \ell(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}), \quad \text{for all } x$$

Lower bound property of LINEARIZATION: Function $\ell(x)$ defined above called **linearization of $f(x)$ at $\bar{x}$**. For convex functions, linearization at any point, lower bound for function at every point.

So approximating convex function by linearization, leads to underestimating it at every point.

How to check if a given function is convex?

One variable functions like $x_1^{\mathbf{2}}, x_1^{\mathbf{4}}, e^{-\mathbf{x_1}}, e^{\mathbf{x_1}}, -\log(x_1)$ [over $x_1 > 0$] are convex.

For many variables, checking convexity hard, as it involves checking PSD of Hessian **at every point !**

Local Convexity: If Hessian of $f(x)$ at $\bar{x}$ is PD, in small neighborhood of $\bar{x}$, $f(x)$ convex. In this case we say $f(x)$ is locally convex at $\bar{x}$.

Types of NLPs

Unconstrained Minimization: No constraints.

Real world problems have constraints. Unconstrained min. very imp. because constrained problems can be transformed into unconstrained ones by penalty function methods.

LP: If all functions affine.

QP: If objective function quadratic, all constraints affine.

Linearly constrained NLP: Objective function nonlinear, all constraints affine.

Equality constrained NLP: All constraints equations, and no bounds on variables.

Convex Programming Problem: Of form

$$
\begin{aligned}
\text{Minimize} \quad & \theta(x) \\
\text{s. to} \quad h_i(x) &= 0, \quad i = 1 \text{ to } m \\
g_p(x) &\geqq 0, \quad p = 1 \text{ to } t
\end{aligned}
$$

where $\theta(x)$ convex, all $h_i(x)$ affine, all $g_p(x)$ concave.

Nicest among NLPs. Useful necessary and sufficient optimality conditions for global minimum are only known for convex programming problems.

Nonconvex Programming Problems: Violates some of the conditions for convex programming.

Types of Solutions For NLP

Feasible solution $\bar{x}$ is

Local Minimum	For $\epsilon > 0$, $\theta(x) \geq \theta(\bar{x})$ for all feasible x satisfying $\|\|x - \bar{x}\|\| < \epsilon$.
Strong local min	if $\theta(x) > \theta(\bar{x})$ for all feasible $x \neq \bar{x}$ satisfying $\|\|x - \bar{x}\|\| < \epsilon$.
Weak local min	If not strong.
Global min	If $\theta(x) \geq \theta(\bar{x})$ for all feasible x.
Stationary Point	If it satisfies necessary condition for local minimum.

Local (strong, weak) maxima, and global maxima similar.

Differences In Constructing LP Models & NLP Models

Variety of functional forms: In LP all affine. In NLP unlimited variety of functional forms.

Data: LP involving m constraints in n variables has $(m+1)(n+1)-1$ coefficients as data elements.

Large scale LP refers to one with $m >$ 1000s, and $n >$ 10,000s. Such models solved to global optimality, in reasonable time.

To construct an NLP model need to determine functional forms for objective, constraint functions. Usually involves Curve Fitting and Parameter estimation. Usually by Least Squares Method using special unconstrained min algos. So even constructing NLP model, needs NLP algos.

So, even 200 variable NLP model considered large scale.

Expectation on solution: LP we don't even talk about local min, because we get global min.

In NLP, if model nonconvex, no efficient algos can guarantee finding global min. So, one compromises on type of solution expected.

Convex Programs Are Nicest NLPs !

Theorem: For convex program every local min is global min.

For convex program, any method finding local min will find global min. Also, every stationary point is global min in convex programs.

Properties of PSD, PD matrices, Algos to check.

Preserved under symmetrization: M PD, PSD, iff $D = \frac{1}{2}(M + M^T)$ is.

Signs of diagonal elements: $M = (m_{ij})$. If PD, all $m_{ii} > 0$. If PSD all $m_{ii} \geq 0$.

Skew-symmetry on a 0 diagonal element: If $M = (m_{ij})$ PSD and $m_{ii} = 0$, then for all j, $m_{ij} + m_{ji} = 0$.

So in symmetric PSD matrix if diagonal element is 0, its row and col must be 0.

Preservation in Principal submatrix after Gaussian Pivot step: Let $D = (d_{ij})$ be symmetric $n \times n$, and suppose $d_{11} \neq 0$.

Perform Gaussian pivot step with d_{11} as pivot element. After pivot step eliminate row 1 and column 1, resulting in matrix D_1 of order $(n-1) \times (n-1)$.

D is PD iff $d_{11} > 0$ and D_1 is PD.

D is PSD iff $d_{11} > 0$ and D_1 is PSD.

Superdiagonalization Algorithm for checking if M PD: **Symmetrize:** Let $D = \frac{1}{2}(M + M^T)$.

Perform Gaussian Pivot Steps: Let $D_0 = D$

Do for $i = 1$ to $n - 1$.

Let D_{i-1} be matrix from previous operation.

If any diagonal entries in $D_{i-1} \leq 0$, terminate. Conclude M not PD.

Otherwise, carry out Gaussian pivot step on D_{i-1} with its ith diagonal element as pivot element, resulting in matrix D_i.

If all Gaussian pivot steps carried out and all diagonal elements of final matrix D_{n-1} are > 0, M PD, terminate.

Examples

$$\begin{pmatrix} 3 & 1 & 2 & 2 \\ -1 & 2 & 0 & 2 \\ 0 & 4 & 4 & \frac{5}{3} \\ 0 & -2 & -\frac{13}{3} & 6 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 2 & 4 & 4 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}$$

Superdiagonalization Algo for checking M PSD

Symmetrize: Let $D = \frac{1}{2}(M + M^T)$.

Perform Gaussian Pivot Steps: Let $E_0 = D$

Do for $i = 1$ to $n - 1$.

Let E_{i-1} be current matrix after previous operation.

If any diagonal entries in E_{i-1} are < 0, terminate. Conclude M not PSD.

If top diagonal entry in E_{i-1} is $= 0$

If top row or 1st col of E_{i-1} are nonzero, terminate. Conclude M not PSD.

Otherwise strike off the zero top row and 1st col of E_{i-1} and let remaining matrix be new current matrix E_i.

If top diagonal entry in E_{i-1} is > 0

Perform Gaussian pivot step on E_{i-1} with the top diagonal element as pivot element.

After this pivot step, erase top row and 1st col of resulting matrix, and let remaining matrix be the new current matrix E_i.

If no termination in above steps, M PSD, terminate.

Examples

$$\begin{pmatrix} 0 & -2 & -3 & -4 & 5 \\ 2 & 3 & 3 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 0 & 0 & 8 & 4 \\ -5 & 0 & 0 & 4 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 2 & 4 & 4 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}$$

Other Properties of PSD, PD Matrices

Property preserved for Principal submatrices: If M PD (PSD) so are all the principal submatrices of M

Sign of determinants: If M is PD (PSD), whether symmetric or not, all its principal subdeterminants are > 0 (≥ 0).

A square symmetric matrix PD iff all its principal subdeterminants are > 0

A square symmetric matrix of order n PD iff all its n *staircase or leading principal subdeterminants* are > 0.

P-matrix: A square matrix, whether symmetric or not, is P-matrix, iff all its principal subdeterminants are > 0.

A symmetric P-matrix is PD. An assymetric P-matrix may not be PD, it may be PSD, or even indefinite.

Linear Dependence Relation at Optimum: If M is PSD, and $\bar{x}$ minimizes $x^T M x$, then $(M + M^T)\bar{x} = 0$.