

Lecture Notes – Part 9

Constrained Nonlinear Programming

1 Introduction

In this subject, we consider the constrained NLP with the linear constraints only. This simplifies the discussion greatly, but does not lose much generality. Most methods for solving optimisation problems with nonlinear constraints transform the constraints into linearised ones, in one way or another, even though it could fail.

2 Linear-Equality Constrained NLP

Consider a minimisation NLP problem with only linear equality constraints:

$$\begin{aligned} \min \quad & z = f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b} \end{aligned} \tag{1}$$

where $\mathbf{A}$ is an $m \times n$ matrix, $n > m$, $\text{rank}(\mathbf{A}) = m$, $\mathbf{x} \in \mathbf{R}^n$, and $\mathbf{b} \in \mathbf{R}^m$. We also assume that the nonlinear function $f(\mathbf{x})$ is defined and has continuous second-order partial derivatives over the feasible region $\mathcal{D} = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}\}$.

Since $f(\mathbf{x})$ is a nonlinear function, its minimum may occur anywhere within the feasible region $\mathcal{D}$ instead of the extreme points. Hence we cannot use any technique similar to the Simplex Method.

2.1 Reduced Function

One of the proposed solution approaches is to transform the considered constrained NLP to an unconstrained NLP by virtue of a *reduced function*.

Example 1. Consider the following problem

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 + x_2^2 \\ \text{s.t.} \quad & 3x_1 + 2x_2 = 6 \end{aligned}$$

We can transform it to an unconstrained version by rearranging the constraint

$$x_2 = 3 - \frac{3}{2}x_1,$$

and substituting it for x_2 in the objective function $f(x_1, x_2)$. Then the considered constrained NLP problem with two decision variables is reduced to an equivalent unconstrained NLP problem with one decision variable

$$\min \quad \phi(x_1) = x_1^2 + \left(3 - \frac{3}{2}x_1\right)^2.$$

Finding the stationary point and applying the optimality conditions for unconstrained NLP, we have a local minimum $(x_1^*, x_2^*)^T = (\frac{18}{13}, \frac{12}{13})$ with the objective value $\min f(x_1^*, x_2^*) = \frac{36}{13}$.

In general case, we transform the constrained NLP (1) to an unconstrained version by means of a general solution to the system of linear equations

$$\mathbf{Ax} = \mathbf{b}. \tag{2}$$

Since the system of constraints (2) is the same as what we have in LP, the similar skill can be utilised. We first divide the n -dimensional decision-variable vector $\mathbf{x}$ into an m -dimensional basis $\mathbf{x}_B$ and an $(n - m)$ -dimensional nonbasis $\mathbf{x}_N$, and correspondingly the $m \times n$ constraint matrix $\mathbf{A}$ into an $m \times m$ basic matrix $\mathbf{B}$ and an $m \times (n - m)$ nonbasic matrix $\mathbf{N}$. Assume without loss of generality that the first m columns of $\mathbf{A}$ give the basic matrix $\mathbf{B}$. Then the vector equation (2) can be rewritten in the form

$$\mathbf{B}\mathbf{x}_B + \mathbf{N}\mathbf{x}_N = \mathbf{b}.$$

So, the basis can be represented in terms of the nonbasis as follows

$$\mathbf{x}_B = \mathbf{B}^{-1}(\mathbf{b} - \mathbf{N}\mathbf{x}_N) = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x}_N.$$

Hence, the general solution of (2) is

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x}_N \\ \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{pmatrix} \mathbf{x}_N, \quad (3)$$

where

$$\begin{pmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{pmatrix} = \bar{\mathbf{x}} \quad \text{and} \quad \begin{pmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{pmatrix} = \mathbf{Z}.$$

Then the general solution (3) can be rewritten in the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{x}_N, \quad (4)$$

where $\bar{\mathbf{x}}$ is a particular solution of (2) by setting $\mathbf{x}_N = \mathbf{0}$, i.e. $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$.

The matrix $\mathbf{Z}$ is called a *basis matrix* for the null-space of $\mathbf{A}$.*

*The null-space of $\mathbf{A}_{m \times n}$ is the set of all $\mathbf{x} \in \mathbf{R}^n$ giving $\mathbf{A}\mathbf{x} = \mathbf{0}$, which is a $(n - m)$ -dimensional subspace of vector space $\mathbf{R}^n$. The set of columns of $\mathbf{Z}$ gives a basis for the null-space of $\mathbf{A}$.

By replacing (4) into the objective function $f(\mathbf{x})$, the constrained NLP (1) is transformed into the unconstrained NLP

$$\min z = f(\bar{\mathbf{x}} + \mathbf{Z}\mathbf{x}_N) = \phi(\mathbf{x}_N)$$

Function $\phi(\mathbf{x}_N)$ of $(n - m)$ -dimensional vector $\mathbf{x}_N$ is called the *reduced function* of the constrained NLP problem (1).

Example 2. Consider the following constrained NLP

$$\begin{aligned} \max \quad & f(x_1, x_2, x_3) = x_1^2 + 4x_1x_3 - x_2^2 \\ \text{s.t.} \quad & 2x_1 + x_2 + 4x_3 = 5 \quad (\text{i}) \\ & 3x_1 + x_2 - x_3 = 1 \quad (\text{ii}) \end{aligned}$$

The feasible set of the problem is the set of solutions of the system of linear equations (i) and (ii).

Applying Gauss-Jordan elimination, we obtain a general solution in the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -4 \\ 13 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ -14 \\ 1 \end{pmatrix} x_3,$$

if we select x_3 as the variable on which x_1 and x_2 are dependent. This is equivalent to selecting $\mathbf{x}_N = x_3$ and $\mathbf{x}_B = (x_1, x_2)^T$, which gives matrices

$$\begin{aligned} \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{B}^{-1} = \begin{pmatrix} -1 & 1 \\ 3 & -2 \end{pmatrix}, \\ \mathbf{B}^{-1}\mathbf{b} = \begin{pmatrix} -1 & 1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} -4 \\ 13 \end{pmatrix} \quad \Rightarrow \quad \bar{\mathbf{x}} = \begin{pmatrix} -4 \\ 13 \\ 0 \end{pmatrix}, \end{aligned}$$

and

$$\mathbf{N} = \begin{pmatrix} 4 \\ -1 \end{pmatrix} \Rightarrow \mathbf{B}^{-1}\mathbf{N} = \begin{pmatrix} -5 \\ 14 \end{pmatrix} \Rightarrow \mathbf{Z} = \begin{pmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} 5 \\ -14 \\ 1 \end{pmatrix}.$$

Hence, the feasible set of the problem is the set of vectors in the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}x_3 = \begin{pmatrix} -4 \\ 13 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ -14 \\ 1 \end{pmatrix} x_3,$$

which is indeed identical to the set obtained by solving the system of equations directly.

Hence the reduced function of the problem is

$$\begin{aligned} \phi(x_3) &= f((-4, 13, 0)^T + (5, -14, 1)^T x_3) = f(5x_3 - 4, -14x_3 + 13, x_3) \\ &= (5x_3 - 4)^2 + 4(5x_3 - 4)x_3 - (-14x_3 + 13)^2. \end{aligned}$$

Again, we can solve the unconstrained NLP problem $\max \phi(x_3)$ by finding the stationary point and applying the optimality conditions for unconstrained NLP. The local maximum is $(x_1^*, x_2^*, x_3^*) = (\frac{166}{151}, -\frac{193}{151}, \frac{154}{151})$ with the objective value $\max f(x_1^*, x_2^*, x_3^*) = \frac{613}{151}$.

2.2 Optimality Conditions

We aim to solve the unconstrained NLP problem with a reduced function

$$\min \phi(\mathbf{x}_\mathbf{N}),$$

where $\phi(\mathbf{x}_\mathbf{N}) = f(\bar{\mathbf{x}} + \mathbf{Z}\mathbf{x}_\mathbf{N})$, by setting up some optimality necessary conditions and sufficient conditions.

Applying the chain rule for differentiation gives the gradient of $\phi(\mathbf{x}_N)$, which is also called the *reduced gradient* of $f(\mathbf{x})$

$$\nabla\phi(\mathbf{x}_N) = \mathbf{Z}^T \nabla f(\mathbf{x}),$$

and the Hessian matrix of $\phi(\mathbf{x}_N)$, which is also called the *reduced Hessian matrix* of $f(\mathbf{x})$

$$\nabla^2\phi(\mathbf{x}_N) = \mathbf{Z}^T \nabla^2 f(\mathbf{x}) \mathbf{Z}.$$

Then Theorems 5 and 6 in Lecture Note – Part 8 imply the following results.

Theorem 1. (Second-order necessary conditions – Linear equality constraints)

If $\mathbf{x}^*$ a local minimiser of $f(\mathbf{x})$ over the set $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$, and $\mathbf{Z}$ is a basis matrix for the null-space of $\mathbf{A}$, then

- (i) $\mathbf{Z}^T \nabla f(\mathbf{x}^*) = \mathbf{0}$, and
- (ii) $\mathbf{Z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{Z}$ is positive semidefinite.

Theorem 2. (Second-order sufficient conditions – Linear equality constraints)

If $\mathbf{Z}$ is a basic matrix for the null-space of $\mathbf{A}$ and the point $\mathbf{x}^*$ satisfies

- (i) $\mathbf{A}\mathbf{x}^* = \mathbf{b}$,
- (ii) $\mathbf{Z}^T \nabla f(\mathbf{x}^*) = \mathbf{0}$, and
- (iii) $\mathbf{Z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{Z}$ is positive definite,

then $\mathbf{x}^*$ a local minimiser of $f(\mathbf{x})$ over the set $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$.

Notice that given a point x for a considered linear-equality constrained NLP problem we can apply directly the above two theorems without deriving a reduced function.

Example 3. Consider the problem

$$\begin{aligned} \min \quad & f(x_1, x_2, x_3) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{s.t.} \quad & x_1 - x_2 + 2x_3 = 2 \end{aligned}$$

We have

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 - 2 \\ 2x_2 \\ -2x_3 + 4 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

To find a basis matrix $\mathbf{Z}$ for the null-space of $\mathbf{A}_{1 \times 3} = (1, -1, 2)$, we choose, for example, $\mathbf{x}_\mathbf{N} = (x_2, x_3)^T$ and $\mathbf{x}_\mathbf{B} = x_1$. Thus, we have $\mathbf{N} = (-1, 2)$ and $\mathbf{B} = 1$. Hence, we have

$$\mathbf{Z} = \begin{pmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we can find the stationary point(s) as per condition (i) and (ii) in Theorem 2. Coupling $x_1 - x_2 + 2x_3 = 2$ with

$$\mathbf{Z}^T \nabla f(\mathbf{x}) = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2x_1 - 2 \\ 2x_2 \\ -2x_3 + 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

we have a system of three equations as follows:

$$\begin{cases} x_1 & -x_2 & +2x_3 & = & 2 \\ 2x_1 & +2x_2 & & = & 2 \\ -4x_1 & & -2x_3 & = & -8. \end{cases}$$

Solving the above system of three equations yields the solution $\mathbf{x}^* = (2.5, -1.5, -1)^T$. Then we check condition (iii) as follows:

$$\mathbf{Z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -4 \\ -4 & 6 \end{pmatrix}.$$

Since the eigenvalues of the Hessian matrix $\mathbf{Z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{Z}$ are $\mu = 5 \pm \sqrt{17} > 0$, $\mathbf{Z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{Z}$ is positive definite.[†] The second-order sufficient conditions are satisfied, so $\mathbf{x}^* = (2.5, -1.5, -1)^T$ is a local minimiser of $f(\mathbf{x})$ over the feasible region.

2.3 The Lagrangian Function

Another technique to transform a constrained NLP to an unconstrained NLP is to introduce the *Lagrangian function*. Consider a minimisation constrained NLP problem

$$\begin{aligned} \min \quad & z = f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) = b_i, \quad i = 1, \dots, m. \end{aligned} \tag{5}$$

We introduce a function called the Lagrangian, which is designed by associating a multiplier λ_i called *Lagrange multiplier*[‡] with the

[†]Notice that the Hessian matrix $\nabla^2 f(\mathbf{x})$ or $\nabla^2 f(\mathbf{x}^*)$ itself is not positive definite.

[‡]The Lagrange multipliers are closely related to the dual variables in duality theorem.

i -th constraint, $i = 1, \dots, m$ in (5), as shown below.

$$\mathbf{L}(\mathbf{x}, \Lambda) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i (b_i - g_i(\mathbf{x})) \quad (6)$$

Then assume that we can find a point $(\mathbf{x}^*, \Lambda^*)$, where $\Lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*) \in \mathbf{R}^m$, so as to minimise $\mathbf{L}(\mathbf{x}, \Lambda)$. Now we explain why $\mathbf{x}^*$ solves (5). If $(\mathbf{x}^*, \Lambda^*)$ minimises $\mathbf{L}(\mathbf{x}, \Lambda)$, then at $(\mathbf{x}^*, \Lambda^*)$

$$\frac{\partial \mathbf{L}(\mathbf{x}, \Lambda)}{\partial \lambda_i} = b_i - g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m. \quad (7)$$

This shows that $\mathbf{x}^*$ will satisfy the constraints in (5) and thus be feasible. To show $\mathbf{x}^*$ is the optimal solution of (5), we let $\mathbf{x}'$ be any feasible solution to (5). Since $(\mathbf{x}^*, \Lambda^*)$ minimises $\mathbf{L}(\mathbf{x}, \Lambda)$, for any vector Λ' we have

$$\mathbf{L}(\mathbf{x}^*, \Lambda^*) \leq \mathbf{L}(\mathbf{x}', \Lambda'). \quad (8)$$

Since $\mathbf{x}^*$ and $\mathbf{x}'$ are both feasible in (5), the terms in (6) involving the λ 's are all zeros, and (8) becomes $f(\mathbf{x}^*) \leq f(\mathbf{x}')$. Thus $\mathbf{x}^*$ does solve (5). In short, if $(\mathbf{x}^*, \Lambda^*)$ solves the unconstrained NLP problem

$$\min \mathbf{L}(\mathbf{x}, \Lambda), \quad (9)$$

then $\mathbf{x}^*$ solves the constrained NLP problem (5).

From Theorem 5 (the first-order necessary condition) in Lecture Note – Part 8, we know that for $(\mathbf{x}^*, \Lambda^*)$ solving (9), it is necessary that at $(\mathbf{x}^*, \Lambda^*)$,

$$\nabla \mathbf{L}(\mathbf{x}, \Lambda) = \mathbf{0}, \quad (10)$$

i.e., $\nabla_{\mathbf{x}}\mathbf{L}(\mathbf{x}, \Lambda) = \mathbf{0}$ and $\nabla_{\Lambda}\mathbf{L}(\mathbf{x}, \Lambda) = \mathbf{0}$. The system of equations $\nabla_{\Lambda}\mathbf{L}(\mathbf{x}, \Lambda) = \mathbf{0}$ is exactly (7), and $\nabla_{\mathbf{x}}\mathbf{L}(\mathbf{x}, \Lambda) = \mathbf{0}$ indicates

$$\nabla f(\mathbf{x}) = \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}),$$

which involves a geometrical interpretation of Lagrange multipliers. For $\mathbf{x}^*$ to solve (5), it is necessary that $\nabla f(\mathbf{x})$ is a linear combination of the constraint gradients $\nabla g_i(\mathbf{x}), i = 1, \dots, m$.

Actually, the linearity of $g_i(\mathbf{x})$ is unnecessary for applying the Lagrangian function. Any point $(\mathbf{x}', \Lambda')$ satisfying (10) is a stationary point for the function $\mathbf{L}(\mathbf{x}, \Lambda)$ and $\mathbf{x}'$ is a feasible solution to (5). That is, we could get the local minimum $\mathbf{x}^*$ to (5) by finding $\mathbf{x}'$ with the smallest objective value among those stationary points.

Example 4. Consider the following constrained NLP

$$\begin{aligned} \min f(x_1, x_2) &= x_1^2 + 2x_2^2 \\ \text{s.t. } x_1^2 + x_2^2 &= 1. \end{aligned}$$

The Lagrangian of the constrained NLP is

$$\mathbf{L}(x_1, x_2, \lambda) = x_1^2 + 2x_2^2 + \lambda(1 - x_1^2 - x_2^2).$$

Then we have

$$\nabla \mathbf{L}(x_1, x_2, \lambda) = \mathbf{0} \quad \Rightarrow \quad \begin{cases} 2x_1 - 2\lambda x_1 = 0 & (\star) \\ 4x_2 - 2\lambda x_2 = 0 & (\star\star) \\ 1 - x_1^2 - x_2^2 = 0 & (\star\star\star) \end{cases}$$

From $(\star)$, we have $x_1 = 0$, $\lambda = 1$ or $\lambda = 0$. But $\lambda = 0$ doesn't satisfy $(\star)$ – $(\star \star \star)$ simultaneously. If $x_1 = 0$, then $(\star \star \star)$ gives $x_2 = 1$ or -1 . If $\lambda = 1$, then we have $x_2 = 0$ from $(\star \star)$, so then $(\star \star \star)$ gives $x_1 = 1$ or -1 . Therefore, we have four stationary points $(x_1, x_2, \lambda) = (0, 1, 2)$, $(0, -1, 2)$, $(1, 0, 1)$, and $(-1, 0, 1)$. Evaluating function $f(x_1, x_2)$ at these four points, we find that $f(0, 1) = 2$, $f(0, -1) = 2$, $f(1, 0) = 1$, and $f(-1, 0) = 1$. Thus, there are two local minima $(x_1^*, x_2^*) = (1, 0)$ and $(-1, 0)$.

If each $g_i(\mathbf{x})$ is a linear function, then we have the following theorem.

Theorem 3. If $f(\mathbf{x})$ is a convex function, then any point $(\mathbf{x}^*, \Lambda^*)$ satisfying (10) will yield a local minimum $\mathbf{x}^*$ to (5).[§]

But if $f(\mathbf{x})$ is not a convex function, then we still have to use (iii) in Theorem 2 to check whether the yielded stationary point(s) are the local minima within the feasible region.

Example 5. Consider the problem in Example 3. The Lagrangian of the constrained NLP is

$$\mathbf{L}(x_1, x_2, x_3, \lambda) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 + \lambda(2 - x_1 + x_2 - 2x_3).$$

Then we have

$$\nabla \mathbf{L}(x_1, x_2, x_3, \lambda) = \mathbf{0} \quad \Rightarrow \quad \begin{cases} 2x_1 - 2 - \lambda = 0 & (11) \\ 2x_2 + \lambda = 0 & (12) \\ -2x_3 + 4 - 2\lambda = 0 & (13) \\ 2 - x_1 + x_2 - 2x_3 = 0 & (14) \end{cases}$$

[§]Note that even if the hypotheses fail to hold, it is still possible that any point satisfying (10) will solve (5).

From (11), (12), and (13) we have

$$\left\{ \begin{array}{lcl} x_1 & = & \frac{\lambda+2}{2} \quad (15) \\ x_2 & = & -\frac{\lambda}{2} \quad (16) \\ x_3 & = & -\lambda + 2 \quad (17) \end{array} \right.$$

Substituting (15), (16) and (17) into (14) gives

$$2 - \frac{\lambda+2}{2} - \frac{\lambda}{2} - 2(-\lambda+2) = \lambda - 3 = 0 \quad \Rightarrow \quad \lambda = 3.$$

Bringing this value into (15), (16) and (17) gives us the solution

$$(x_1, x_2, x_3)^T = (2.5, -1.5, -1),$$

which is the unique stationary point. Since we have seen that the second-order sufficient conditions are satisfied at $(x_1, x_2, x_3)^T = (2.5, -1.5, -1)$ in Example 3, this is the local minimum over the feasible region.