

# The general nonlinear program

Mark S. Gockenbach

## 1 Introduction

The general nonlinear programming problem has the form

$$\min f(x) \tag{1}$$

$$s.t. \quad g(x) = 0 \tag{2}$$

$$h(x) \geq 0, \tag{3}$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ . Optimality conditions for (1–3) are obtained by combining the optimality conditions for equality-constrained and inequality-constrained problems in a straightforward fashion. Similarly, sequential unconstrained minimization techniques for the general NLP can be devised by combining the logarithmic barrier method with either the quadratic penalty method or the augmented Lagrangian method. Because the results are straightforward extensions of the previous material, I will leave most proofs as exercises.

## 2 Optimality conditions

The first step in deriving the optimality conditions is to determine the correct constraint qualification and notion of regularity. I will continue to use  $\mathcal{A}(x)$  to denote the (indices of the) set of active *inequality* constraints for a feasible point  $x$ :

$$\mathcal{A}(x) = \{i : 1 \leq i \leq p, h_i(x) = 0\}.$$

Given a feasible point  $x^*$  for (1–2), the vector  $z$  should define a feasible direction if it satisfies

$$\nabla g(x^*)^T z = 0, \tag{4}$$

$$\nabla h_i(x^*) \cdot z \geq 0, \quad i \in \mathcal{A}(x^*). \tag{5}$$

However, unless further conditions are satisfied, there may exist vectors  $z$  satisfying the above conditions and yet not tangent to any feasible path through  $x^*$ . The constraint qualification is therefore

**Constraint qualification 2.1** *For all  $z \in \mathbb{R}^n$  satisfying (4–5), there exists a feasible path  $x : [0, a] \rightarrow \mathbb{R}^n$  such that*

$$x(0) = x^*,$$

$$\dot{x}(0) = z.$$

Moreover,  $x$  can be chosen so that

$$i \in \mathcal{A}(x^*), \nabla h_i(x^*) \cdot z = 0 \Rightarrow h(x(t)) = 0 \text{ for all } t \in [0, a].$$

The appropriate notion of regularity is the following:

**Definition 2.2** A feasible point  $x^*$  is a regular point of (1-3) if

$$\{\nabla g_i(x^*) : i = 1, 2, \dots, m\} \cup \{\nabla y_i(x^*) : i \in \mathcal{A}(x^*)\}$$

is linearly independent.

As usual, regularity implies the constraint qualification:

**Theorem 2.3** Suppose  $x^*$  is a regular point of NLP (1-3), where  $f$ ,  $g$ , and  $h$  are smooth. Then  $x^*$  satisfies constraint qualification 2.1.

The proof is very similar to the proof given in the case of equality constraints only.

The following theorem gives the first-order necessary conditions for the general NLP. The reader will notice the  $\lambda \in \mathbb{R}^m$  denotes the Lagrange multiplier corresponding to the equality constraints and  $\theta \in \mathbb{R}^p$  the Lagrange multiplier corresponding to the inequality constraints.

**Theorem 2.4** Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are continuously differentiable and  $x^*$  is a local minimizer of the NLP

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

If the constraint qualification holds at  $x^*$  (in particular, if  $x^*$  is a regular point of the NLP), then there exist  $\lambda^* \in \mathbb{R}^m$  and  $\theta^* \in \mathbb{R}^p$  such that the following conditions hold:

$$\begin{aligned} \nabla f(x^*) &= \nabla g(x^*)\lambda^* + \nabla h(x^*)\theta^*, \\ g(x^*) &= 0, \\ h(x^*) &\geq 0, \\ \theta^* &\geq 0, \\ \theta_i^* h_i(x^*) &= 0, \quad i = 1, 2, \dots, p. \end{aligned}$$

**Proof:** Since the constraint qualification holds by assumption, if  $z \in \mathbb{R}^n$  satisfies

$$\begin{aligned} \nabla g_i(x^*) \cdot z &= 0, \quad i = 1, 2, \dots, m, \\ \nabla h_i(x^*) \cdot z &\geq 0, \quad i \in \mathcal{A}(x^*), \end{aligned}$$

there exists a feasible path  $x : [0, a] \rightarrow \mathbb{R}^n$  such that  $x(0) = x^*$  and  $\dot{x}(0) = z$ . Defining  $\phi : [0, a] \rightarrow \mathbb{R}$  by

$$\phi(t) = f(x(t)),$$

it follows that  $\phi$  has a local minimizer at  $t = 0$ , and hence that  $\phi'(0) \geq 0$ , that is,

$$\nabla f(x^*) \cdot z \geq 0.$$

I have therefore proved

$$\left. \begin{aligned} \nabla g_i(x^*) \cdot z &= 0, \quad i = 1, 2, \dots, m, \\ \nabla h_i(x^*) \cdot z &\geq 0, \quad i \in \mathcal{A}(x^*) \end{aligned} \right\} \Rightarrow \nabla f(x^*) \cdot z \geq 0. \quad (6)$$

I need to use a slight trick in order to apply Farkas's lemma. The reader will notice that

$$\nabla g_i(x^*) = 0$$

is equivalent to the two inequalities

$$\nabla g_i(x^*) \geq 0, \quad -\nabla g_i(x^*) \geq 0.$$

I can therefore write (6) as

$$\left. \begin{array}{l} \nabla g_i(x^*) \cdot z \geq 0, \quad i = 1, 2, \dots, m, \\ -\nabla g_i(x^*) \cdot z \geq 0, \quad i = 1, 2, \dots, m, \\ \nabla h_i(x^*) \cdot z \geq 0, \quad i \in \mathcal{A}(x^*) \end{array} \right\} \Rightarrow \nabla f(x^*) \cdot z \geq 0. \quad (7)$$

Farkas's lemma then implies the existence of vectors  $\lambda^+ \in \mathbb{R}^m$ ,  $\lambda^- \in \mathbb{R}^m$ ,  $\theta^* \in \mathbb{R}^p$  such that

$$\lambda^+ \geq 0, \quad \lambda^- \geq 0, \quad \theta^* \geq 0, \quad \theta_i^* = 0 \text{ if } i \notin \mathcal{A}(x^*)$$

and

$$\begin{aligned} \nabla f(x^*) &= \sum_{i=1}^m \lambda_i^+ \nabla g_i(x^*) - \sum_{i=1}^m \lambda_i^- \nabla g_i(x^*) + \sum_{i=1}^p \theta_i^* \nabla h_i(x^*) \\ &= \sum_{i=1}^m (\lambda_i^+ - \lambda_i^-) \nabla g_i(x^*) + \sum_{i=1}^p \theta_i^* \nabla h_i(x^*) \\ &= \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^p \theta_i^* \nabla h_i(x^*), \\ &= \nabla g(x^*) \lambda^* + \nabla h(x^*) \theta^*, \end{aligned}$$

where  $\lambda^* = \lambda^+ - \lambda^-$ . (The reader will notice that although  $\lambda^+ \geq 0$  and  $\lambda^- \geq 0$ , the components of  $\lambda^*$  can be positive, negative, or zero.) QED

The Lagrangian of (1–3) is defined to be the function  $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$  defined by

$$\ell(x; \lambda; \theta) = f(x) - \lambda \cdot g(x) - \theta \cdot h(x).$$

The Lagrange multiplier condition then states that  $x^*$  is a stationary point of  $\ell(\cdot; \lambda^*; \theta^*)$ :

$$\nabla \ell(x^*; \lambda^*; \theta^*) = 0.$$

**Example 2.5** Consider the NLP

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) = 0 \\ & h(x) \geq 0, \end{array}$$

where

$$\begin{aligned} f(x) &= (x_1 - 1)^2 + 2(x_2 + 2)^2 + 3(x_3 + 3)^2, \\ g(x) &= x_3 - x_2 - x_1 - 1, \\ h(x) &= x_3 - x_1^2 \end{aligned}$$

(three independent variables, one equality constraint and one inequality constraint). It is easy to see that this is a convex program, so any local solution is a global solution. The gradients are given by

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} 2x_1 - 2 \\ 4x_2 + 8 \\ 6x_3 + 18 \end{bmatrix}, \\ \nabla g(x) &= \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \\ \nabla h(x) &= \begin{bmatrix} -2x_1 \\ 0 \\ 1 \end{bmatrix}, \end{aligned}$$

and so the first-order necessary conditions form the system of equations

$$\begin{aligned} 2x_1 - 2 &= -\lambda - 2\theta x_1, \\ 4x_2 + 8 &= -\lambda, \\ 6x_3 + 18 &= \lambda + \theta, \\ x_3 - x_2 - x_1 &= 1, \\ \theta(x_3 - x_1^2) &= 0 \end{aligned}$$

together with the inequalities

$$\begin{aligned} x_3 - x_1^2 &\geq 0, \\ \theta &\geq 0. \end{aligned}$$

Using MATLAB's Symbolic Toolbox, I find that the system of equations has two real solutions. However, one of the solutions corresponds to an infeasible point ( $h(x) < 0$ ) and so the unique  $(x, \lambda, \theta)$  satisfying the first-order conditions is

$$\begin{aligned} x^* &= (0.12288, -1.1078, 0.015100), \\ \lambda^* &= -3.5689, \\ \theta^* &= 21.659. \end{aligned}$$

Since the NLP is a convex program, this defines the unique global solution of the problem.

Second-order optimality conditions are derived as before. I will write  $D(x^*, \theta^*)$  for the set of  $z \in \mathbb{R}^n$  satisfying the following conditions:

$$\begin{aligned} \nabla g_i(x^*) \cdot z &= 0, \quad i = 1, 2, \dots, m, \\ \nabla h_i(x^*) \cdot z &= 0, \quad i \in \mathcal{A}(x^*), \theta_i^* > 0, \\ \nabla h_i(x^*) \cdot z &\geq 0, \quad i \in \mathcal{A}(x^*), \theta_i^* = 0. \end{aligned}$$

**Theorem 2.6** Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are continuously differentiable and  $x^*$  is a local minimizer of the NLP

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

Suppose further that the constraint qualification holds at  $x^*$  (for example,  $x^*$  is a regular point of the NLP), and that  $\lambda^* \in \mathbb{R}^m$  and  $\theta^* \in \mathbb{R}^p$  are Lagrange multipliers corresponding to the equality and inequality constraints, respectively. Then

$$z \cdot \nabla^2 \ell(x^*; \lambda^*; \theta^*) z \geq 0 \text{ for all } z \in D(x^*, \theta^*).$$

Sufficient conditions for a local minimizer are given in the following theorem:

**Theorem 2.7** Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are twice continuously differentiable and  $x^* \in \mathbb{R}^n$ ,  $\lambda^* \in \mathbb{R}^m$ ,  $\theta^* \in \mathbb{R}^p$  satisfy the following conditions:

$$\begin{aligned} \nabla f(x^*) &= \nabla g(x^*)\lambda^* + \nabla h(x^*)\theta^*, \\ g(x^*) &= 0, \\ h(x^*) &\geq 0, \\ \theta^* &\geq 0, \\ \theta_i^* h_i(x^*) &= 0, \quad i = 1, 2, \dots, p, \\ z \cdot \nabla^2 \ell(x^*; \lambda^*; \theta^*) z &> 0 \text{ for all } z \in D(x^*, \theta^*), z \neq 0. \end{aligned}$$

Then  $x^*$  is a strict local minimizer of the NLP

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

### 3 SUMT for the general NLP

#### 3.1 Using the logarithmic barrier method

One way to use successive unconstrained minimization to solve the general NLP (1–3) is to combine the quadratic penalty function or augmented Lagrangian with the logarithmic barrier function. I will focus on the augmented Lagrangian and leave the (simpler) quadratic penalty function to the reader.

I define the functional

$$G(x; \lambda; \mu; \eta) = f(x) - \lambda \cdot g(x) + \frac{1}{2\mu} \|g(x)\|^2 - \eta \sum_{i=1}^p \log(h_i(x)).$$

One could use a single penalty parameter  $\mu$  (replacing  $\eta$  by  $\mu$  in the above formula). However,  $\mu$  need not go to zero in the augmented Lagrangian method, while driving  $\eta$  to zero is mandatory in the logarithmic barrier technique. Therefore, it is probably better to allow the two parameters to be adjusted separately. The algorithm then takes the following form:

Choose  $x^{(0)}$ ,  $\mu_0$ ,  $\lambda^{(0)}$ , and  $\eta_0$  such that  $h(x^{(0)}) > 0$ ,  $\mu_0 > 0$ , and  $\eta_0 > 0$ .

For  $k = 1, 2, 3, \dots$

Using  $x^{(k-1)}$  as a starting point, solve

$$\min_x G(x; \lambda^{(k-1)}; \mu_{k-1}; \eta_{k-1})$$

to get  $x^{(k)}$ .

Set  $\lambda^{(k)} = \lambda^{(k-1)} - \mu_{k-1}^{-1} g(x^{(k-1)})$ .

Choose  $\mu_k \in (0, \mu_{k-1})$  and  $\eta_k \in (0, \eta_{k-1}]$ .

The gradient of  $G$  is given by

$$\begin{aligned} J(x; \lambda; \mu; \eta) &= \nabla f(x) - \nabla g(x) \lambda + \frac{1}{\mu} \nabla g(x) g(x) - \eta \sum_{i=1}^p \frac{1}{h_i(x)} \nabla h_i(x) \\ &= \nabla f(x) - \nabla g(x) (\lambda - \mu^{-1} g(x)) - \nabla h(x) (\eta h(x)^{-1}) \\ &= \nabla f(x) - \nabla g(x) \lambda_+ - \nabla h(x) \theta, \end{aligned}$$

where

$$\begin{aligned} \lambda_+ &= \lambda - \mu^{-1} g(x), \\ \theta &= \eta h(x)^{-1}, \end{aligned}$$

or, equivalently,

$$\begin{aligned} -g(x) - \mu(\lambda_+ - \lambda) &= 0, \\ \theta h(x) - \eta e &= 0. \end{aligned}$$

Therefore, the equation  $\nabla G(x; \lambda; \mu; \eta) = 0$  can be written as the following system of  $n + m + p$  equations for the  $n + m + p$  unknowns  $x, \lambda_+, \theta$ :

$$\begin{aligned}\nabla f(x) - \nabla g(x)\lambda_+ - \nabla h(x)\theta &= 0, \\ -g(x) - \mu(\lambda_+ - \lambda) &= 0, \\ \theta h(x) - \eta e &= 0.\end{aligned}$$

This system is a perturbation of the following equations that form part of the first-order necessary conditions:

$$\begin{aligned}\nabla f(x) - \nabla g(x)\lambda - \nabla h(x)\theta &= 0, \\ -g(x) &= 0, \\ \theta h(x) &= 0.\end{aligned}$$

By applying the implicit function theorem in the usual way, one can prove that, for all  $\mu, \eta$  sufficiently small,  $G$  has unique local minimizer  $x(\lambda, \mu, \eta)$  in a neighborhood of a given nonsingular point  $x^*$ , and that  $x(\lambda, \mu, \eta) \rightarrow x^*$  as  $\lambda \rightarrow \lambda^*$  and  $\eta \rightarrow 0$ . Moreover, one can prove that updating  $\lambda$  by the formula  $\lambda_+ = \lambda - \mu^{-1}g(x(\lambda, \mu, \eta))$  leads to causes the Lagrange multiplier  $\lambda$  estimate to converge to  $\lambda^*$ , even if  $\mu$  is bounded away from zero.

**Example 3.1** *I solve the NLP from Example (2.5) using the algorithm described above. I take*

$$x^{(0)} = (0, 0, 1), \quad \lambda^{(0)} = 0, \quad \mu_0 = 1, \quad \eta_0 = 1$$

and define

$$\mu_k = \frac{1}{4^k}, \quad \eta_k = \frac{1}{10^k}.$$

The algorithm continued until

$$\|g(x^{(k)})\| < 10^{-8}, \quad \theta^{(k)}h(x^{(k)}) < 10^{-8}.$$

The result was 10 iterations, with

$$\|g(x^{(10)})\| \doteq 6.2172 \cdot 10^{-15}, \quad \theta^{(10)}h(x^{(10)}) \doteq 1.0000 \cdot 10^{-9}$$

and

$$\begin{aligned}x^{(10)} &\doteq (0.12288, -1.1078, 0.015100), \\ \lambda^{(10)} &\doteq -3.5689, \\ \theta^{(10)} &\doteq 21.662.\end{aligned}$$

Comparing to the results in Example 2.5 shows that only  $\theta^{(10)}$  is in error to the precision displayed above. This is due to the relatively slow convergence of the logarithmic barrier method as compared to the augmented Lagrangian method.

### 3.2 The augmented Lagrangian method for general NLPs

Another sequential unconstrained minimization method for general NLPs uses the augmented Lagrangian method to handle both equality and inequality constraints, using slack variables to convert the inequality constraints to equations. As described earlier, the slack variables are eliminated analytically from the problem, so that the unconstrained minimizations are performed with respect to the original variables  $x$  only.

Adding slack variables, the NLP

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

is transformed to

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) - s = 0 \\ & s \geq 0. \end{aligned}$$

The augmented Lagrangian is then

$$L(x, s; \lambda; \theta; \mu; \eta) = f(x) - \lambda \cdot g(x) + \frac{1}{2\mu} \|g(x)\|^2 - \theta \cdot (h(x) - s) + \frac{1}{2\eta} \|h(x) - s\|^2$$

and

$$\begin{aligned} \tilde{L}(x; \lambda; \theta; \mu; \eta) &= \min \{L(x, s; \lambda; \theta; \mu; \eta) : s \geq 0\} \\ &= f(x) - \lambda \cdot g(x) + \frac{1}{2\mu} \|g(x)\|^2 + \sum_{i=1}^p \Phi(h_i(x), \theta_i, \eta). \end{aligned}$$

The function  $\Phi$  is defined as before:

$$\Phi(\alpha; \beta; \gamma) = \begin{cases} \frac{1}{2\gamma} \alpha^2 - \beta\alpha, & \alpha \leq \beta\gamma, \\ -\frac{1}{2}\gamma\beta^2, & \alpha > \beta\gamma. \end{cases}$$

The result is the following algorithm:

Choose  $x^{(0)}$ ,  $\lambda^{(0)}$ ,  $\mu_0 > 0$ , and  $\eta_0 > 0$ .

For  $k = 1, 2, 3, \dots$

Using  $x^{(k-1)}$  as a starting point, solve

$$\min_x \tilde{L}(x; \lambda^{(k-1)}; \theta^{(k-1)}; \mu_{k-1}; \eta_{k-1})$$

to get  $x^{(k)}$ .

Set  $\lambda^{(k)} = \lambda^{(k-1)} - \mu_{k-1}^{-1} g(x^{(k-1)})$ .

Set  $\theta^{(k)} = \max \{ \theta^{(k-1)} - \eta_{k-1}^{-1} h(x^{(k-1)}), 0 \}$ .

Choose  $\mu_k \in (0, \mu_{k-1}]$  and  $\eta \in (0, \eta_{k-1}]$ .

**Example 3.2** I now solve the NLP from Example (2.5) using the augmented Lagrangian algorithm. I take

$$x^{(0)} = (0, 0, 1), \lambda^{(0)} = 0, \theta^{(0)} = 0, \mu_0 = 1, \eta_0 = 1$$

and reduce both  $\mu$  and  $\eta$  by a factor of 10 at each iteration (that is,  $\mu_k = \eta_k = 10^{-k}$ ). The algorithm continued until

$$\|g(x^{(k)})\| < 10^{-8}, \theta^{(k)} h(x^{(k)}) < 10^{-8}.$$

The result was 6 iterations, with

$$\|g(x^{(6)})\| \doteq 6.7042 \cdot 10^{-11}, \theta^{(6)} h(x^{(6)}) \doteq 2.992 \cdot 10^{-9}$$

and

$$\begin{aligned}x^{(6)} &\doteq (0.12288, -1.1078, 0.015100), \\ \lambda^{(6)} &\doteq -3.5689, \\ \theta^{(6)} &\doteq 21.659.\end{aligned}$$

Comparing to the results in Example 2.5 shows that the computed solution is correct to the five digits displayed (actually, each of  $x^{(6)}$ ,  $\lambda^{(6)}$ , and  $\theta^{(6)}$  is accurate to 7–10 digits).