1. Constrained Optimization

We will present optimality conditions, followed by a brief outline of ideas behind algorithms for constrained optimization.

OPTIMALITY CONDITIONS

We start by considering problems involving equality constraints only, and then move on to discussing the more general case of problems with both equality and inequality constraints.

1.1. Problems with equality constraints

We consider a problem with equality constraints in the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array}$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}$, $h(x) = [h_1(x), h_2(x), \ldots, h_m(x)]^T : \mathbb{R}^n \to \mathbb{R}^m$. We assume that $m < n$. Denote by

$$J_h(x) = \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \frac{\partial h_1(x)}{\partial x_2} & \cdots & \frac{\partial h_1(x)}{\partial x_n} \\ \frac{\partial h_2(x)}{\partial x_1} & \frac{\partial h_2(x)}{\partial x_2} & \cdots & \frac{\partial h_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m(x)}{\partial x_1} & \frac{\partial h_m(x)}{\partial x_2} & \cdots & \frac{\partial h_m(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla h_1(x)^T \\ \nabla h_2(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix}$$

the Jacobian of h at x. Let $X = \{x \in \mathbb{R}^n : h(x) = 0\}$ denote the feasible set of the considered optimization problem. We will call $x^* \in \mathbb{R}^n$ a regular point if $J_h(x^*)$ has the full rank, that is, $\text{rank}(J_h(x^*)) = m$, or equivalently, $\nabla h_1(x^*), \nabla h_2(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent.

We first discuss the FONC for the case with two variables and a single constraint: $n = 2$ and $m = 1$. We can write such a problem as follows:

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) \\ \text{subject to} & h_1(x_1, x_2) = 0. \end{array}$$

Then the feasible set $X = \{x \in \mathbb{R}^2 : h_1(x_1, x_2) = 0\}$ is the level set of h_1 at the level 0.

Let $x^* = [x_1^*, x_2^*]^T$ be a regular point and a local minimizer of this problem. Then $h_1(x^*) = 0$, and $\nabla h_1(x^*)$ is orthogonal to the tangent line to any curve passing through x^* in the level set X. Consider an arbitrary curve $\gamma = \{y(t) = [x_1(t), x_2(t)]^T,\ t \in [\alpha, \beta]\} \subset X$ passing through x^* in X, that is, $y(t^*) = x^*$ for some $t^* \in (\alpha, \beta)$. The direction of the tangent line to γ at x^* is given by $y'(t)$, hence we have

$$\nabla h_1(x^*)^T y'(t^*) = 0.$$

Since x^* is a local minimizer of our problem, it will remain a local minimizer if we restrict the feasible region to the points of the curve γ. Therefore, t^* is a local minimizer of the problem $\min\limits_{t \in [\alpha, \beta]} f(y(t))$, and since t^* is an interior point of $[\alpha, \beta]$, t^* is a local minimizer of the single-variable unconstrained problem

$$\min_{t \in \mathbb{R}} f(y(t)).$$

Using the FONC for this unconstrained problem, we have

$$\frac{df(y(t^*))}{dt} = 0.$$

On the other hand, from the chain rule,

$$\frac{df(y(t^*))}{dt} = \nabla f(y(t^*))^T y'(t^*),$$

so

$$\nabla f(y(t^*))^T y'(t^*) = 0.$$

Thus, $\nabla f(y(t^*))$ is orthogonal to $y'(t^*)$. So, we have shown that if $\nabla f(x^*) \neq 0$, then $\nabla f(x^*)$ and $\nabla h_1(x^*)$ are both orthogonal to the same vector $y'(x^*)$. For 2-dimensional vectors, this means that $\nabla f(x^*)$ and $\nabla h_1(x^*)$ are parallel, implying that there exists a scalar λ such that

$$\nabla f(x^*) + \lambda \nabla h_1(x^*) = 0.$$

A similar property holds for the general case and is formulated in the following theorem.

Theorem **(Lagrange Theorem)** *If x^* is a regular point and a local minimizer (maximizer) of the problem*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array}$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}$, $h(x) = [h_1(x), \ldots, h_m(x)]^T : \mathbb{R}^n \to \mathbb{R}^m$, then there exists $\lambda = [\lambda_1, \ldots, \lambda_m]^T \in \mathbb{R}^m$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0.$$

Here, $\lambda_i, i = 1, \ldots, m$, are called the *Lagrange multipliers* and the function

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

is called the *Lagrangian* of the considered problem. Note that $L(x, \lambda)$ is a function of $n + m$ variables. If we apply the unconstrained FONC to this function we obtain the system

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) &= 0 \\ h(x^*) &= 0, \end{aligned}$$

which coincides with the FONC stated in the Lagrange theorem (the second equation just guarantees the feasibility). This system has $n + m$ variables and $n + m$ equations. Its solutions are candidate points for a local minimizer (maximizer). The system is not easy to solve in general. Moreover, like in the unconstrained case, even if we solve it, a solution may not be a local minimizer–it can be a saddle point or a local maximizer. Figure 1 illustrates the FONC.

Example 1 *Apply the FONC (Lagrange theorem) to the problem*

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & 4x_1^2 + x_2^2 - 1 = 0. \end{array}$$

Note that all feasible points are regular for this problem, so any local minimizer has to satisfy the Lagrange conditions. We have

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda(4x_1^2 + x_2^2 - 1),$$

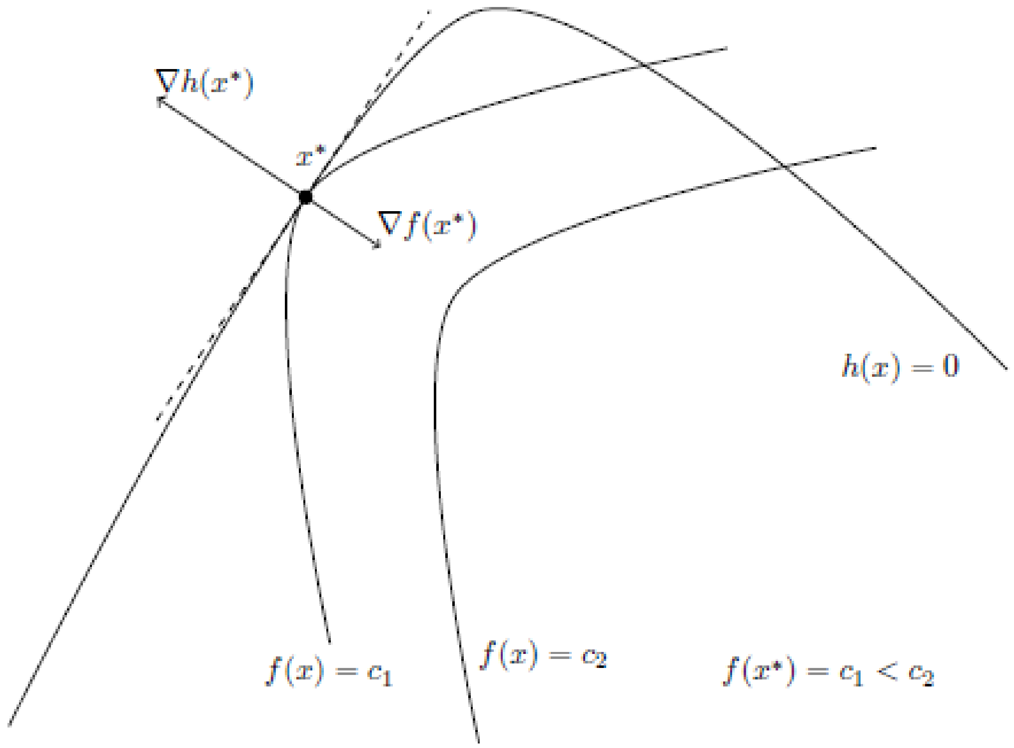

FIGURE 1: An illustration of the FONC for equality-constrained problems. Here x^* satisfies the FONC and is a local maximizer, which is not global.

and the Lagrange conditions give the system

$$\begin{aligned} 2x_1(1+4\lambda) &= 0 \\ 2x_2(1+\lambda) &= 0 \\ 4x_1^2 + x_2^2 &= 1. \end{aligned}$$

This system has four solutions:

(1) $\lambda_1 = -1, x^{(1)} = [0, 1]^T$;
(2) $\lambda_2 = -1, x^{(2)} = [0, -1]^T$;
(3) $\lambda_3 = -1/4, x^{(3)} = [1/2, 0]^T$;
(4) $\lambda_4 = -1/4, x^{(4)} = [-1/2, 0]^T$.

From a geometric illustration (Figure .2), it is easy to see that $x^{(1)}$ and $x^{(2)}$ are global maximizers, whereas $x^{(3)}$ and $x^{(4)}$ are global minimizers.

Example 2 *Consider the problem of optimizing $f(x) = x^T Q x$ subject to a single equality constraint $x^T P x = 1$, where P is a positive definite matrix. Apply the FONC to this problem.*

The Lagrangian is

$$L(x, \lambda) = x^T Q x + \lambda(1 - x^T P x).$$

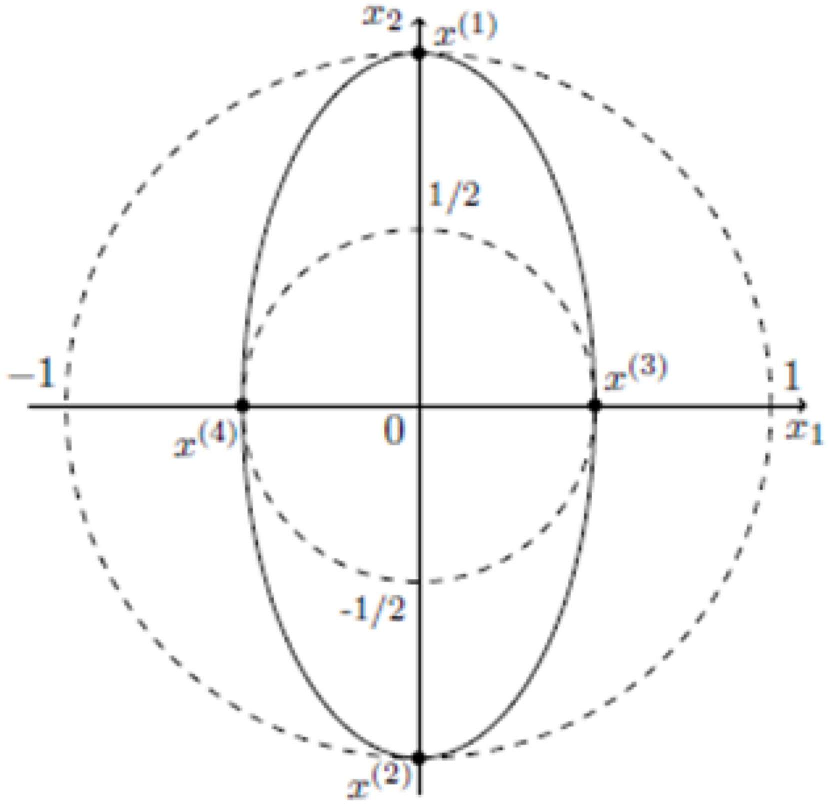

FIGURE 2: Illustration of Example 1.

Convex case

Consider a convex problem with equality constraints,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array}$$

where $f(x)$ is a convex function and $X = \{x \in \mathbb{R}^n : h(x) = 0\}$ is a convex set. We will show that the Lagrange theorem provides sufficient conditions for a global minimizer in this case.

Theorem 2 *Let x^* be a regular point satisfying the Lagrange theorem,*

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

and

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0.$$

Then x^* is a global minimizer.

Proof. From the first-order characterization of a convex function, we have

$$f(x) - f(x^*) \geq \nabla f(x^*)^T (x - x^*), \ \ \forall x \in X.$$

From the FONC

$$\nabla f(x^*) = -\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*).$$

So

$$f(x) - f(x^*) \geq -\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)^T (x - x^*).$$

Note that for any i, $\nabla h_i(x^*)^T(x - x^*)$ is the directional derivative of h_i at x^* in the direction $x - x^*$. Hence, using the definition of the directional derivative we obtain

$$\begin{aligned} \nabla h_i(x^*)^T (x - x^*) &= \lim_{\alpha \to 0+} \frac{h_i(x^* + \alpha(x - x^*)) - h_i(x^*)}{\alpha} \\ &= \lim_{\alpha \to 0+} \frac{h_i(\alpha x + (1 - \alpha) x^*) - h_i(x^*)}{\alpha} \\ &= 0, \end{aligned}$$

since $\alpha x + (1 - \alpha)x^* \in X$ due to the convexity of X and $h(y) = 0$ for any $y \in X$. Substituting this result into we obtain

$$f(x) - f(x^*) \geq 0.$$

Since x is an arbitrary point in X, x^* is a global minimizer of the considered problem. □

1.2. Problems with inequality constraints

We consider the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0, \end{array}$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}$, $h(x) : \mathbb{R}^n \to \mathbb{R}^m$ $(m < n)$, and $g(x) : \mathbb{R}^n \to \mathbb{R}^p$.

This problem involves two types of constraints, equality and inequality constraints. Recall that an inequality constraint $g_j(x) \leq 0$ is called *active* at x^* if $g_j(x^*) = 0$. We denote by $I(x^*) = \{j : g_j(x^*) = 0\}$ the set of indices corresponding to the active constraints for x^*. A point x^* is called a *regular point* for the considered problem if $\nabla h_i(x^*), i = 1, \ldots, m$ and $\nabla g_j(x^*), j \in I(x^*)$ form a set of linearly independent vectors. The Lagrangian of this problem is defined as

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x),$$

where $\lambda = [\lambda_1, \ldots, \lambda_m]^T \in \mathbb{R}^m$ and $\mu = [\mu_1, \ldots, \mu_p]^T \in \mathbb{R}^p, \mu \geq 0$. As before, the multipliers $\lambda_i, i = 1, \ldots, m$ corresponding to the equality constraints are called the Lagrange multipliers. The multipliers $\mu_j, j = 1, \ldots, p$ corresponding to the inequality constraints are called the Karush-Kuhn-Tucker (KKT) multipliers.

The first-order necessary conditions for the problems with inequality constraints are referred to as Karush-Kuhn-Tucker (KKT) conditions.

Theorem 3 (Karush-Kuhn-Tucker (KKT) conditions)
If x^* is a regular point and a local minimizer of the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \textit{subject to} & h(x) = 0 \\ & g(x) \leq 0, \end{array}$$

where all functions are continuously differentiable, then there exist $\lambda \in \mathbb{R}^m$ and a nonnegative $\mu \in \mathbb{R}^p$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0$$

and

$$\mu_j g_j(x^*) = 0, \; j = 1, \ldots, p \;\; \textit{(Complementary slackness).}$$

In summary, the KKT conditions can be expressed in terms of the following system of equations and inequalities:

1. $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$;
2. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0$;
3. $\mu_j g_j(x^*) = 0$, $j = 1, \ldots, p$;
4. $h(x^*) = 0$;
5. $g(x^*) \leq 0$.

Example 4 *Consider the problem*

$$\begin{array}{ll} \textit{minimize} & x_2 - x_1 \\ \textit{subject to} & x_1^2 + x_2^2 \leq 4 \\ & (x_1+1)^2 + x_2^2 \leq 4. \end{array}$$

The Lagrangian is

$$L(x, \mu) = x_2 - x_1 + \mu_1(x_1^2 + x_2^2 - 4) + \mu_2((x_1+1)^2 + x_2^2 - 4).$$

Using the KKT conditions we have the system

$$\begin{aligned} -1 + 2\mu_1 x_1 + 2\mu_2(x_1+1) &= 0 \\ 1 + 2\mu_1 x_2 + 2\mu_2 x_2 &= 0 \\ \mu_1(x_1^2 + x_2^2 - 4) &= 0 \\ \mu_2\left((x_1+1)^2 + x_2^2 - 4\right) &= 0 \\ x_1^2 + x_2^2 &\leq 4 \\ (x_1+1)^2 + x_2^2 &\leq 4 \\ \mu_1, \mu_2 &\geq 0. \end{aligned}$$

1. *If $\mu_1 = 0$, then the system becomes*

$$\begin{aligned} 2\mu_2(x_1+1) &= 1 \\ 2\mu_2 x_2 &= -1 \\ \mu_2\left((x_1+1)^2 + x_2^2 - 4\right) &= 0 \\ x_1^2 + x_2^2 &\leq 4 \\ (x_1+1)^2 + x_2^2 &\leq 4 \\ \mu_2 &\geq 0. \end{aligned}$$

Note that $\mu_2 \neq 0$ (from the first equation), so $\mu_2 > 0$. Adding the first two equations we get

$$x_1 = -x_2 - 1,$$

which, using the third equation, gives

$$x_2 = \pm\sqrt{2}.$$

Since $\mu_2 > 0$*, from the second equation* $x_2 = -\sqrt{2}$ *and* $\mu_2 = 1/(2\sqrt{2})$*, so* $x_1 = \sqrt{2}-1$*. These values of* x_1 *and* x_2 *satisfy the inequality constraints, thus* $x^* = [\sqrt{2}-1, -\sqrt{2}]^T$ *satisfies the KKT conditions.*

2. *If* $\mu_2 = 0$*, then the system becomes*

$$\begin{aligned} 2\mu_1 x_1 &= 1 \\ 2\mu_1 x_2 &= -1 \\ \mu_1(x_1^2 + x_2^2 - 4) &= 0 \\ x_1^2 + x_2^2 &\leq 4 \\ (x_1+1)^2 + x_2^2 &\leq 4 \\ \mu_1 &\geq 0. \end{aligned}$$

Note that $\mu_1 \neq 0$ *(from the first equation), so* $\mu_1 > 0$*. Adding the first two equations we get*

$$x_1 = -x_2,$$

which, using the third equation, gives

$$x_2 = \pm\sqrt{2}.$$

Since $\mu_1 > 0$*, from the second equation* $x_2 = -\sqrt{2}$ *and* $\mu_1 = 1/(2\sqrt{2})$*, so* $x_1 = \sqrt{2}$*. However, these values of* x_1 *and* x_2 *do not satisfy the last inequality constraint, hence the point is infeasible and the KKT conditions are not satisfied.*

3. *If* $\mu_1 \neq 0, \mu_2 \neq 0$*, then the system becomes*

$$\begin{aligned} -1 + 2\mu_1 x_1 + 2\mu_2(x_1+1) &= 0 \\ 1 + 2\mu_1 x_2 + 2\mu_2 x_2 &= 0 \\ x_1^2 + x_2^2 - 4 &= 0 \\ (x_1+1)^2 + x_2^2 - 4 &= 0 \\ \mu_1, \mu_2 &\geq 0. \end{aligned}$$

From the last two equalities we obtain $x_1 = -1/2$*,* $x_2 = \pm\sqrt{15}/2$*. Solving the first two equations with* $x_1 = -1/2$*, for* $x_2 = \sqrt{15}/2$ *we get* $\mu_1 = -\frac{1}{2}(1+1/\sqrt{15}), \mu_2 = \frac{1}{2}(1-1/\sqrt{15})$*, whereas for* $x_2 = -\sqrt{15}/2$ *we obtain* $\mu_1 = -\frac{1}{2}(1-1/\sqrt{15}), \mu_2 = \frac{1}{2}(1+1/\sqrt{15})$*. In both cases, one of the KKT multipliers is negative, so these points do not satisfy the KKT conditions.*

From Figure 3 it is clear that the KKT point x^* *is the global minimizer. The level set of the objective corresponding to the optimal value* $(1-2\sqrt{2})$ *is shown by the dashed line (a tangent to the feasible region).*

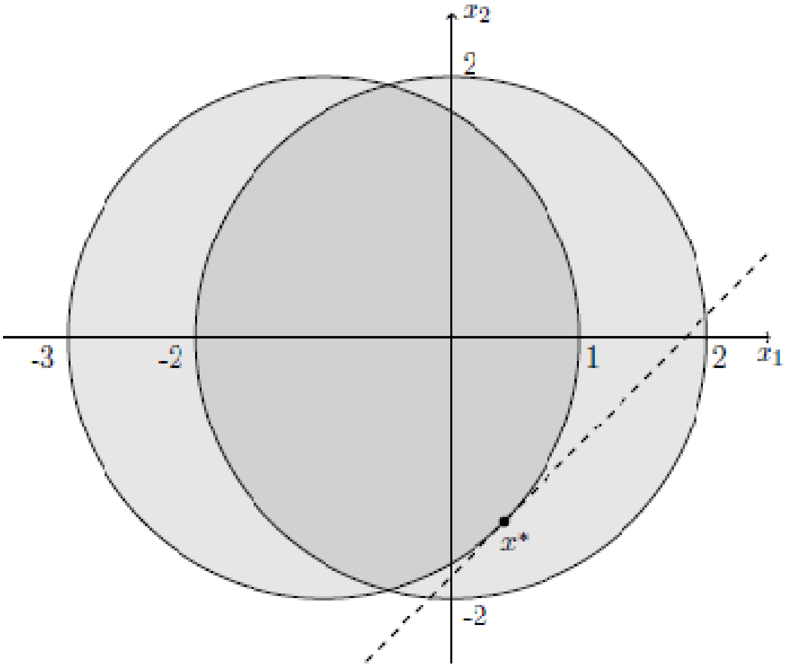

FIGURE 3: An illustration of Example 4.

Convex case

Assume that the feasible set $X = \{x : h(x) = 0, g(x) \le 0\}$ is a convex set and $f(x)$ is a convex function over X. We will show that in this case the KKT conditions are sufficient conditions for a global minimizer.

Theorem 4 *Consider a convex problem*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \textit{subject to} & h(x) = 0 \\ & g(x) \le 0, \end{array}$$

where all functions are continuously differentiable, the feasible set $X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \le 0\}$ *is convex, and* $f(x)$ *is convex on* X*. A regular point* x^* *is a global minimizer of this problem if and only if it satisfies the KKT conditions, that is, there exist* $\lambda \in \mathbb{R}^m$ *and a nonnegative* $\mu \in \mathbb{R}^p$ *such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0$$

and

$$\mu_j g_j(x^*) = 0, \ j = 1, \ldots, p.$$