High-Order Optimality Conditions

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/˜yyye

Optimality Conditions: How to recognize an optimizer

The duality theorem establishes an optimality condition for convex optimization and is called the zero-order condition. Could one explore more structures of the functions in the objective and constraints to construct more concrete and executable conditions?

High-order derivative information: The objective and constraint are often specified by functions that are continuously differentiable or in C^1 over certain regions.

Sometimes the functions are twice continuously differentiable or in C^2 over certain regions.

The theory distinguishes these two cases and develops first-order optimality conditions and second-order optimality conditions.

Observation from One-Variable Problem

Consider a differentiable function f of one variable defined on an interval F. If an interior-point $\bar{x}$ is a local/global minimizer, then $f'(\bar{x}) = 0$. If the left-end-point $\bar{x}$ is a local minimizer, then $f'(\bar{x}) \geq 0$. This is called the first-order necessary condition.

If $f'(\bar{x}) = 0$, then it is necessary that $f(x)$ is a locally convex function at $\bar{x}$, so that $f''(\bar{x}) \geq 0$ is also necessary. This is called the second-order necessary condition.

These conditions are not, in general, sufficient. It does not distinguish between local minimizers, local maximizers, or points of inflection. However, if in addition to the first-order condition, the second-order condition $f''(\bar{x}) > 0$ is satisfied, then $\bar{x}$ is a local minimizer. This is a second-order sufficient condition.

If the function is convex, the first order necessary condition is also sufficient.

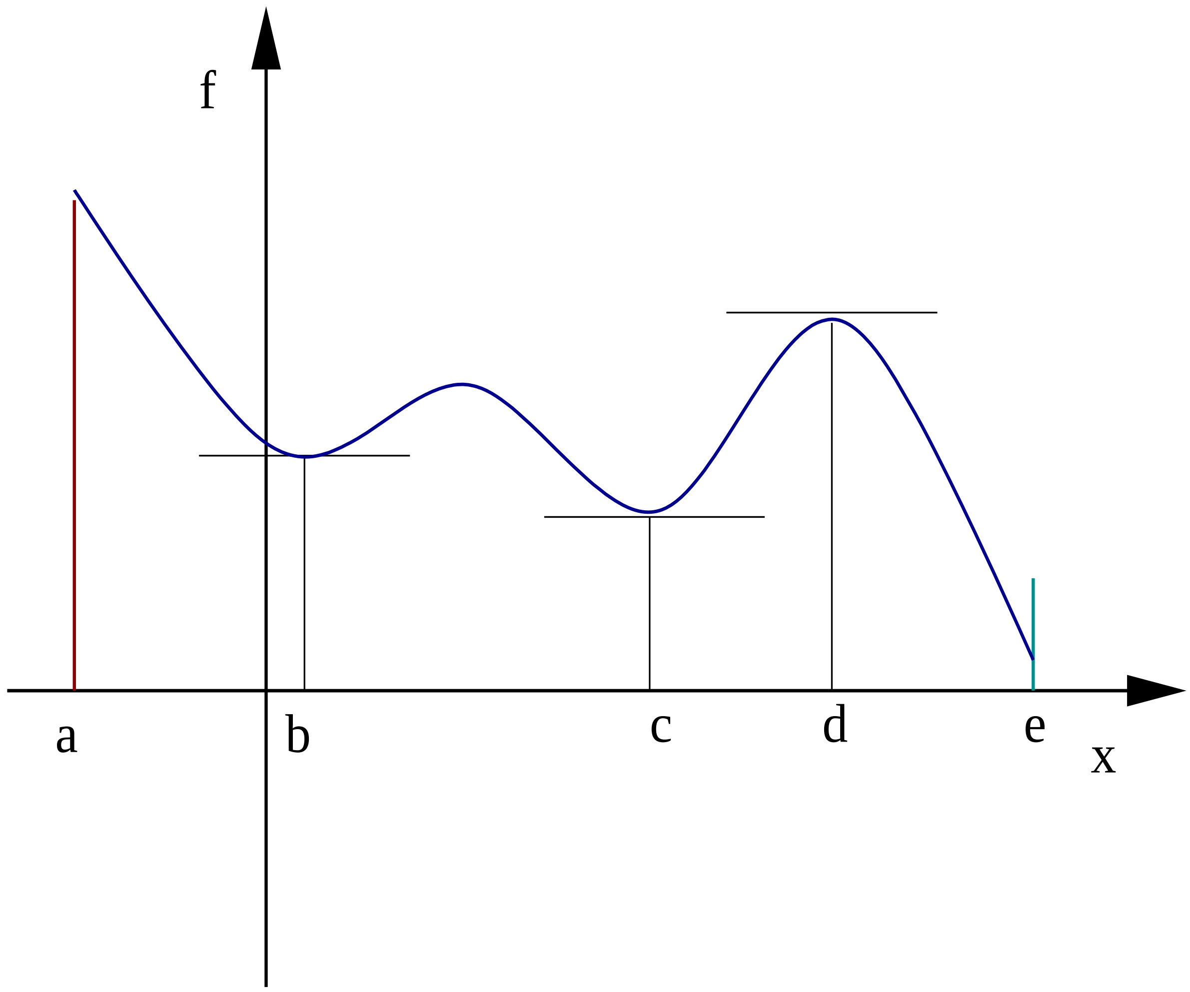

Figure 1: Global and local minimizers of one-variable function

Conditions for Unconstrained Optimization

Theorem 1 *(First-Order Optimality Condition) Let* $f(\mathbf{x})$ *be a* C^1 *function where* $\mathbf{x} \in R^n$*. Then, if* $\mathbf{x}^*$ *is a (local) minimizer, it is necessarily* $\nabla f(\mathbf{x}^*) = \mathbf{0}$*.*

The first-order condition will be sufficient if $f(\mathbf{x})$ is a convex function.

Theorem 2 *(Second-Order Optimality Condition) Let* $f(\mathbf{x})$ *be a* C^2 *function where* $\mathbf{x} \in R^n$*. Then, if* $\mathbf{x}^*$ *is a minimizer, it is necessarily*

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \quad \textit{and} \quad \nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}.$$

Furthermore, if $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$*, then the condition becomes sufficient.*

The proofs would be based on Taylor's theorem such that if these conditions are not satisfied, then one would be find a descent-direction vector $\mathbf{d}$ and a small constant $\bar{\alpha} > 0$ such that $f(\mathbf{x}^* + \alpha\mathbf{d}) < f(\mathbf{x}^*),\ \forall 0 < \alpha \leq \bar{\alpha}$.

First-Order Condition for Convex Optimization I

Consider the constrained problem again: find $\mathbf{x} \in R^n$ to

$$\begin{array}{lll} (COP) & \min & f(\mathbf{x}) \\ & \text{s.t.} & \mathbf{h}(\mathbf{x}) = \mathbf{0} \in R^m, \\ & & \mathbf{c}(\mathbf{x}) \le \mathbf{0} \in R^p. \end{array}$$

Recall Lagrangian Function

$$L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = f(\mathbf{x}) + \mathbf{y}^T \mathbf{h}(\mathbf{x}) + \mathbf{s}^T \mathbf{c}(\mathbf{x}),$$

and Lagrangian Relaxation Problem for given Lagrange multipliers $(\mathbf{y}, \mathbf{s} \ge \mathbf{0})$:

$$(LRP) \quad \min_{\mathbf{x}} \quad L(\mathbf{x}, \mathbf{y}, \mathbf{s}).$$

Under convexity and certain regularity conditions, there are multipliers $(\mathbf{y}, \mathbf{s} \ge \mathbf{0})$ such that the optimizers of (COP) and (LRP) coincide and $s_i c_i(\mathbf{x}) = 0$ for all i.

First-Order Condition for Convex Optimization II

Theorem 3 *(First-Order or KKT Optimality Condition) Let (COP) be a convex minimization problem and let (COP) have an interior-point feasible solution, that is, there is* $\hat{\mathbf{x}}$ *such that* $\mathbf{c}(\hat{\mathbf{x}}) < \mathbf{0}$*. Then, if* $\mathbf{x}^*$ *is a minimizer of (LRP), it is necessarily*

$$\nabla_x L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*) = \mathbf{0}$$

and

$$s_i^* c_i(\mathbf{x}^*) = 0, \ \forall i$$

for some multipliers $(\mathbf{y}^*, \mathbf{s}^* \geq \mathbf{0})$.

Note that (we treat the gradients as row vectors):

$$\nabla_x L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \nabla f(\mathbf{x}) + \mathbf{y}^T \nabla \mathbf{h}(\mathbf{x}) + \mathbf{s}^T \nabla \mathbf{c}(\mathbf{x}).$$

There gradient vectors of all functions involved in (COP) are linearly dependent at $\mathbf{x}^*$.

Linear Programming again

Standard LP case: $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$, $\mathbf{h}(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ and $\mathbf{c}(\mathbf{x}) = -\mathbf{x}$:

$$\nabla_x L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \mathbf{c}^T + \mathbf{y}^T A - \mathbf{s}^T.$$

$$\begin{aligned} x_j s_j &= 0, \ \forall j = 1, \ldots, n \\ A\mathbf{x} &= \mathbf{b} \\ \mathbf{c} + A^T\mathbf{y} - \mathbf{s} &= \mathbf{0} \\ \mathbf{x}, \mathbf{s} &\geq \mathbf{0} \end{aligned}$$

These are identical conditions derived from conic duality.

The objective ball tangents the constraint hyperplane

Consider the problem

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

$$\bar{\mathbf{x}} = \left(\frac{1}{2};\ \frac{1}{2}\right).$$

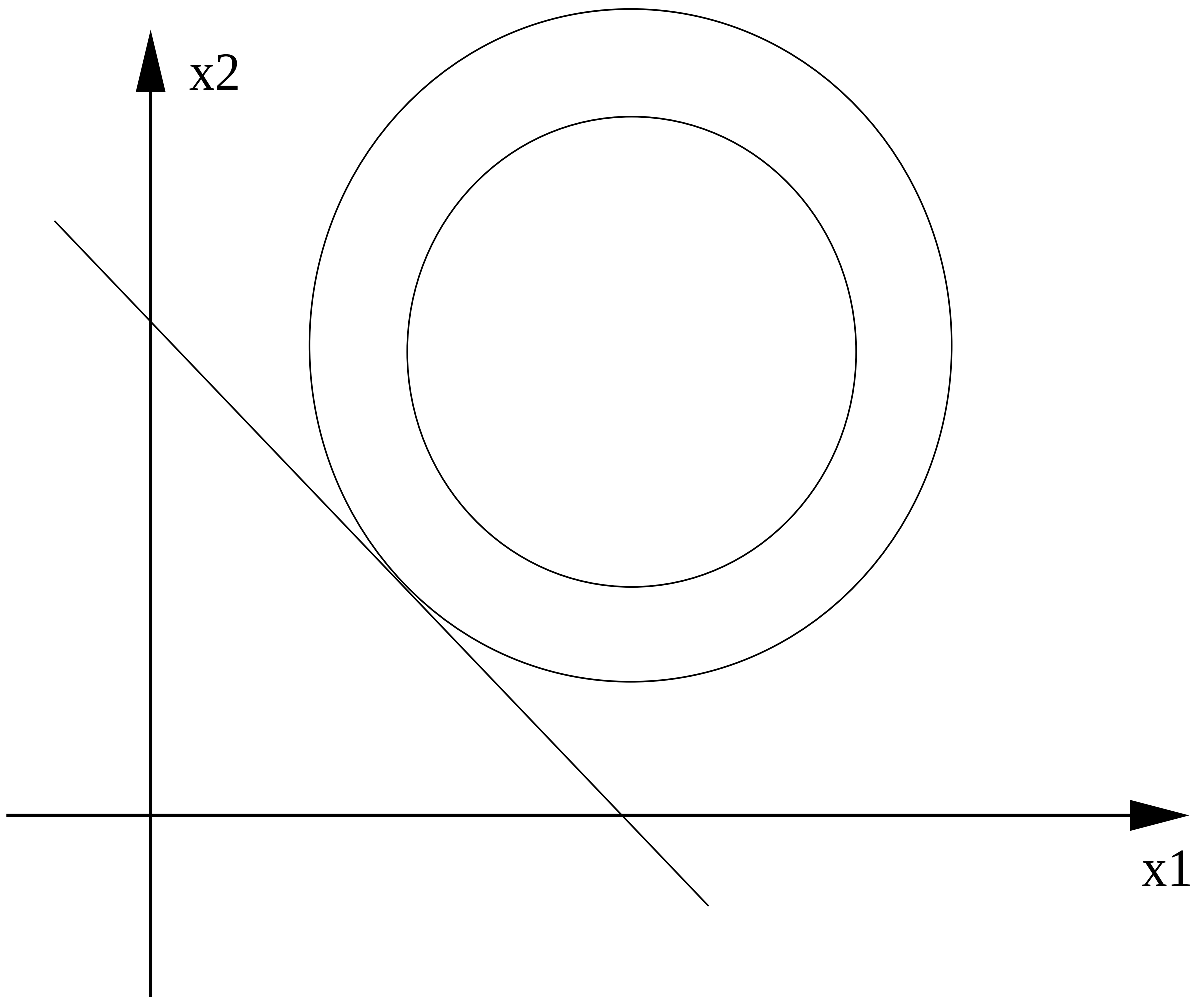

Figure 2: The objective ball tangents the constraint hyperplane

The objective gradient in the normal cone of the constraint hyperplane

Consider the problem

$$\begin{array}{ll} \text{minimize} & (x_1-1)^2+(x_2-1)^2 \\ \text{subject to} & x_1+2x_2-1 \le 0, \\ & 2x_1+x_2-1 \le 0. \end{array}$$

$$\bar{\mathbf{x}} = \left(\frac{1}{3};\ \frac{1}{3}\right).$$

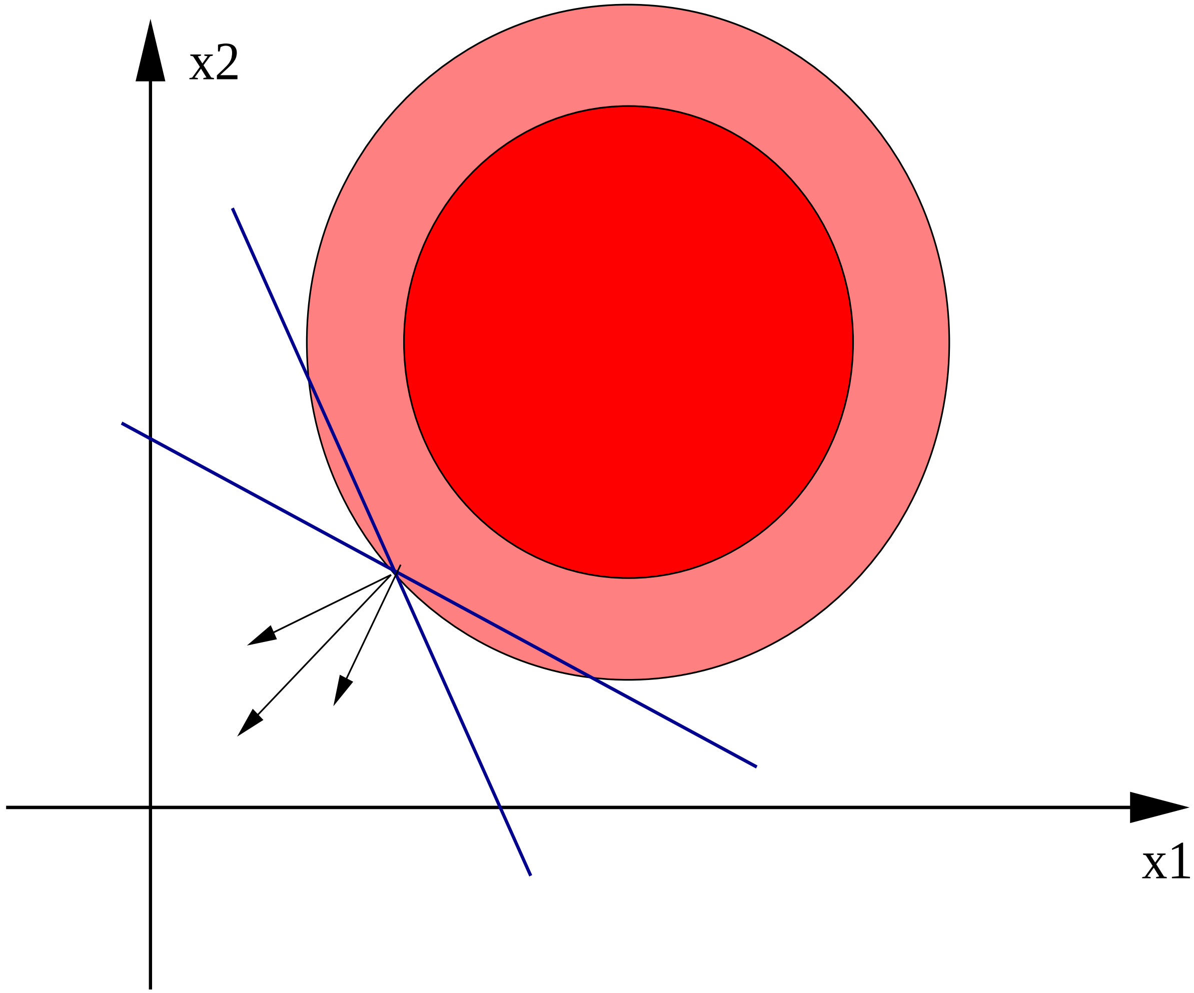

Figure 3: The objective gradient in the normal cone of the constraint hyperplane

The feasible region and objective level set are tangential

Consider the problem

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

$$\bar{\mathbf{x}} = \left(\frac{1}{\sqrt{2}};\ \frac{1}{\sqrt{2}}\right)$$

or

$$\bar{\mathbf{x}} = \left(\frac{-1}{\sqrt{2}};\ \frac{-1}{\sqrt{2}}\right).$$

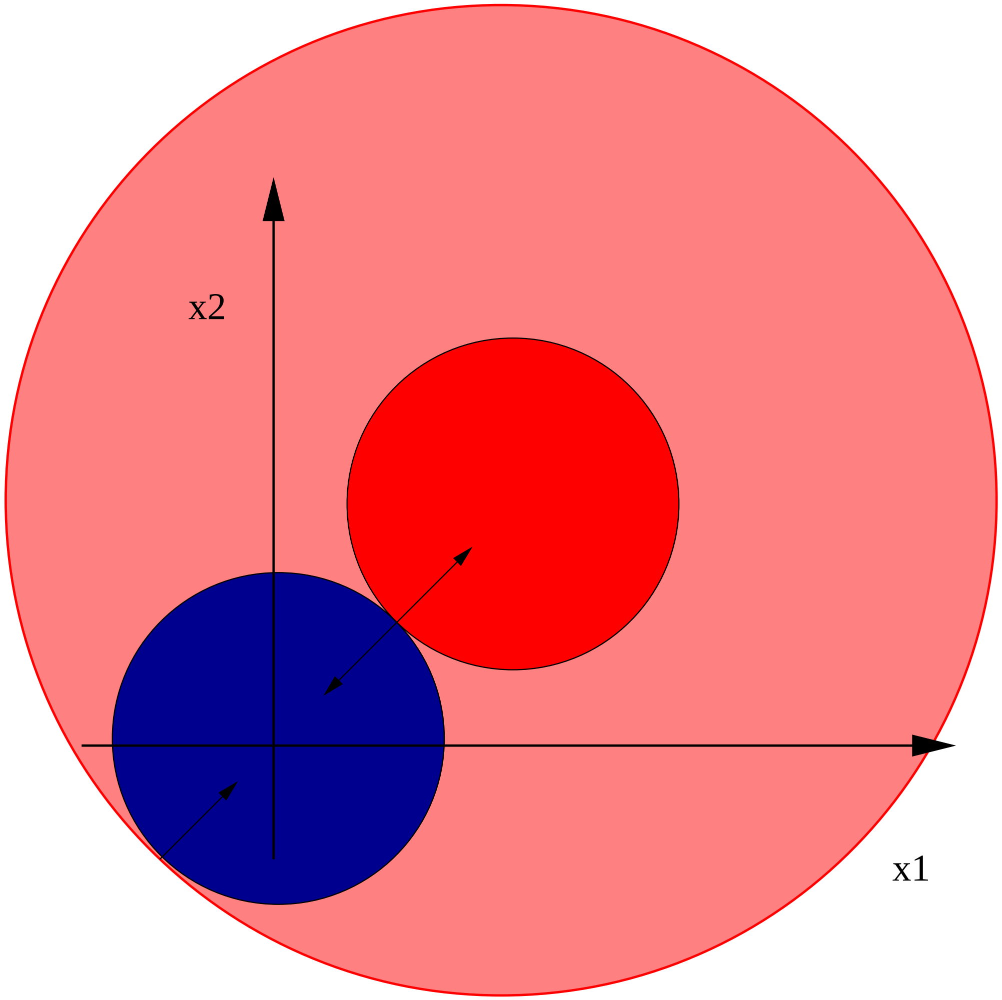

Figure 4: The two spheres tangent to each other at two points, but one has a wrong sign of the multiplier

First-Order Condition is Sufficient for Convex Optimization

Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ be a KKT point for (COP) in which $\bar{\mathbf{x}}$ is a feasible vector. Consider the Lagrangian function $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + \mathbf{s}^T\mathbf{c}(\mathbf{x})$ associated with (COP). Let $\mathbf{x}$ be feasible and $\mathbf{s} \geq \mathbf{0}$. By our hypotheses, L is a convex and differentiable function of $\mathbf{x}$. Hence by the gradient inequality applied to L

$$L(\mathbf{x}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) \geq L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) + \nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})(\mathbf{x} - \bar{\mathbf{x}}) \text{ for all feasible } \mathbf{x}.$$

More explicitly,

$$\begin{aligned} & f(\mathbf{x}) + \bar{\mathbf{y}}^T(A\mathbf{x} - \mathbf{b}) + \bar{\mathbf{s}}^T\mathbf{c}(\mathbf{x}) \\ \geq\ & f(\bar{\mathbf{x}}) + \bar{\mathbf{y}}^T(A\bar{\mathbf{x}} - \mathbf{b}) + \bar{\mathbf{s}}^T\mathbf{c}(\bar{\mathbf{x}}) + \nabla_x L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})(\mathbf{x} - \bar{\mathbf{x}}). \end{aligned}$$

Hence, together with $A\mathbf{x} - \mathbf{b} = A\bar{\mathbf{x}} - \mathbf{b} = \mathbf{0}$, $\bar{\mathbf{s}}^T\mathbf{c}(\bar{\mathbf{x}}) = 0$, and $\nabla_x L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) = \mathbf{0}$, we have $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) - \bar{\mathbf{s}}^T\mathbf{c}(\mathbf{x}) \geq f(\bar{\mathbf{x}})$.

First-Order Condition for Convex-Constrained Nonlinear Optimization I

In the following, we consider cases where the feasible region is convex but the objective is general, called Convex-Constrained Nonlinear Optimization (CCNO). Now, if $\mathbf{x}^*$ is a (local) minimizer of (CCNO), then $\mathbf{x}^*$ must be a minimizer of the following convex-constrained linear optimization problem:

$$\begin{array}{ll} \min & \nabla f(\mathbf{x}^*)\mathbf{x} \\ \text{s.t.} & \mathbf{h}(\mathbf{x}) = \mathbf{0} \in R^m, \\ & \mathbf{c}(\mathbf{x}) \leq \mathbf{0} \in R^p. \end{array}$$

The proofs would be based on contradiction: if the statement is not true, one would find another feasible solution $\bar{\mathbf{x}}$ such that $\nabla f(\mathbf{x}^*)(\bar{\mathbf{x}} - \mathbf{x}^*) < 0$. Let $\mathbf{d} = \bar{\mathbf{x}} - \mathbf{x}^*$. Then $\mathbf{d}$ is a descent-direction vector. From Taylor's theorem there is a small constant $\bar{\alpha} > 0$ such that $f(\mathbf{x}^* + \alpha\mathbf{d}) < f(\mathbf{x}^*),\ \forall 0 < \alpha \leq \bar{\alpha}$ and $\mathbf{x}^* + \alpha\mathbf{d}$ remains feasible for all $0 \leq \alpha \leq \bar{\alpha}$.

The optimizer remains optimal for the linearized

Consider the problem

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

$$\bar{\mathbf{x}} = \left(\frac{-1}{\sqrt{2}};\ \frac{-1}{\sqrt{2}} \right).$$

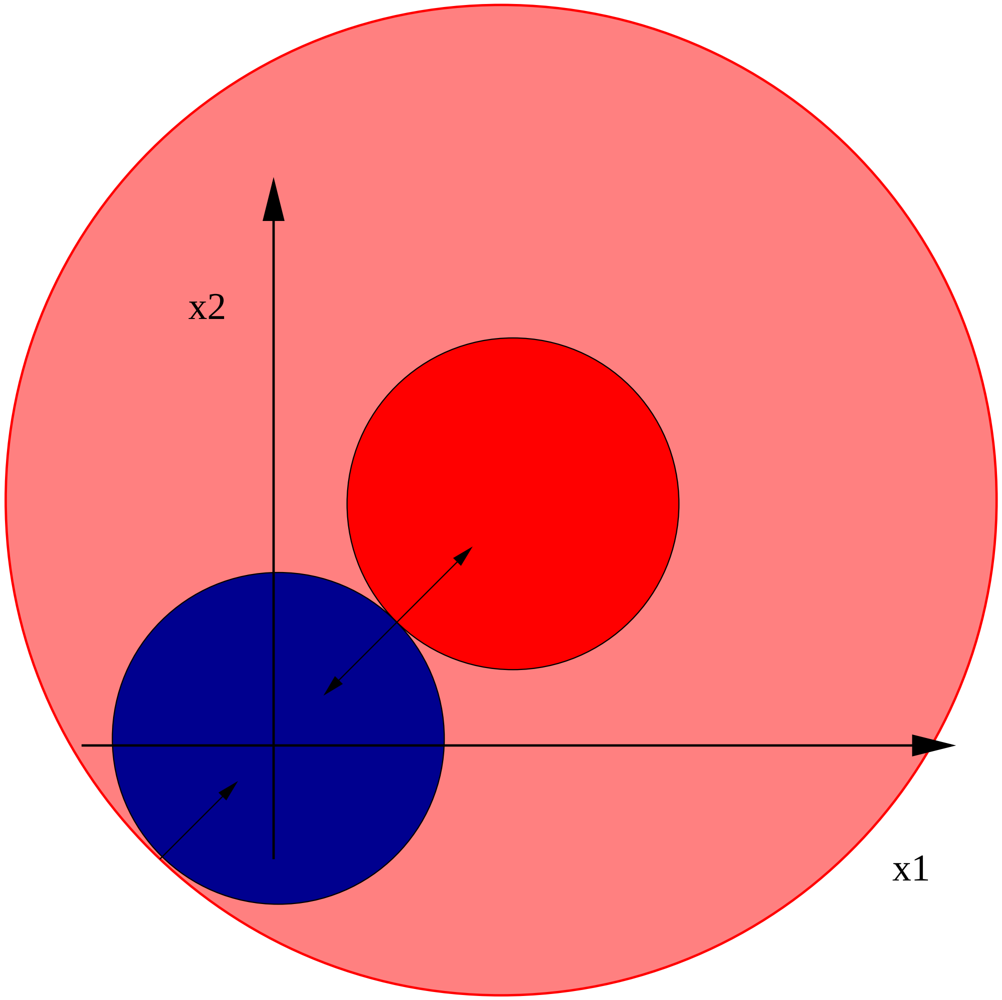

Figure 5: Possible Optimizers of CCNO

First-Order Condition for Convex-Constrained Nonlinear Optimization II

Therefore, the same First-Order first-order (necessary) condition applies:

Theorem 4 *(First-Order or KKT Optimality Condition) Let the feasible region of (CCNO) be convex and have an interior-point feasible solution, that is, there is* $\hat{\mathbf{x}}$ *such that* $\mathbf{c}(\hat{\mathbf{x}}) < \mathbf{0}$*. Then, if* $\mathbf{x}^*$ *is a (local) minimizer, it is necessarily*

$$\nabla_x L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*) = \mathbf{0}$$

and

$$s_i^* c_i(\mathbf{x}^*) = 0, \ \forall i$$

for some multipliers $(\mathbf{y}^*, \mathbf{s}^* \geq \mathbf{0})$.

Second-Order Necessary Conditions for CCNO

Consider CCNO, and in addition assume function f is twice continuously differentiable. Let F denote the feasible region of (CCNO). For a given $\mathbf{x}^* \in F$, we define the active-constraint set $\mathcal{C}^* = \{i: \ c_i(\mathbf{x}^*) = 0\}$. Let

$$T^* := \{\mathbf{z} : \nabla\mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0}, \ \nabla c_i(\mathbf{x}^*)\mathbf{z} = 0 \forall i \in \mathcal{C}^*\}.$$

T^* is sometimes called the tangent linear space of the active constraints at $\mathbf{x}^*$.

Theorem 5 *Let $\mathbf{x}^*$ be a (local) minimizer of (CCNO) and let $\mathbf{y}^*, \mathbf{s}^*$ denote Lagrange multipliers such that $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ satisfies the (first-order) KKT conditions of (CCNO). Then, it is also necessary to have*

$$\mathbf{z}^T \nabla^2_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)\mathbf{z} \geq 0 \qquad \forall \mathbf{z} \in T^*.$$

The proof is based on that fact that $\mathbf{x}^*$ is a local minimizer of the Lagrangian Relaxation Problem on the tangent space so that the Hessian of the Lagrangian function need to be positive semidefinite on the tangent space.

Second-Order Sufficient Conditions for CCNO

Theorem 6 *Let* $\mathbf{x}^*$ *be a feasible solution of (CCNO) and let* $\mathbf{y}^*, \mathbf{s}^*$ *be the Lagrange multipliers such that* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ *satisfies the (first-order) KKT conditions of (CCNO). Then, if in addition*

$$\mathbf{z}^T \nabla^2_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)\mathbf{z} > 0 \qquad \forall \mathbf{0} \neq \mathbf{z} \in T^*,$$

then $\mathbf{x}^*$ *is a local minimizer of (CCNO).*

The proof can be found in Chapter 11.8 of LY.