Optimality Conditions for General Constrained Optimization

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/˜yyye

Optimality Condition Illustration

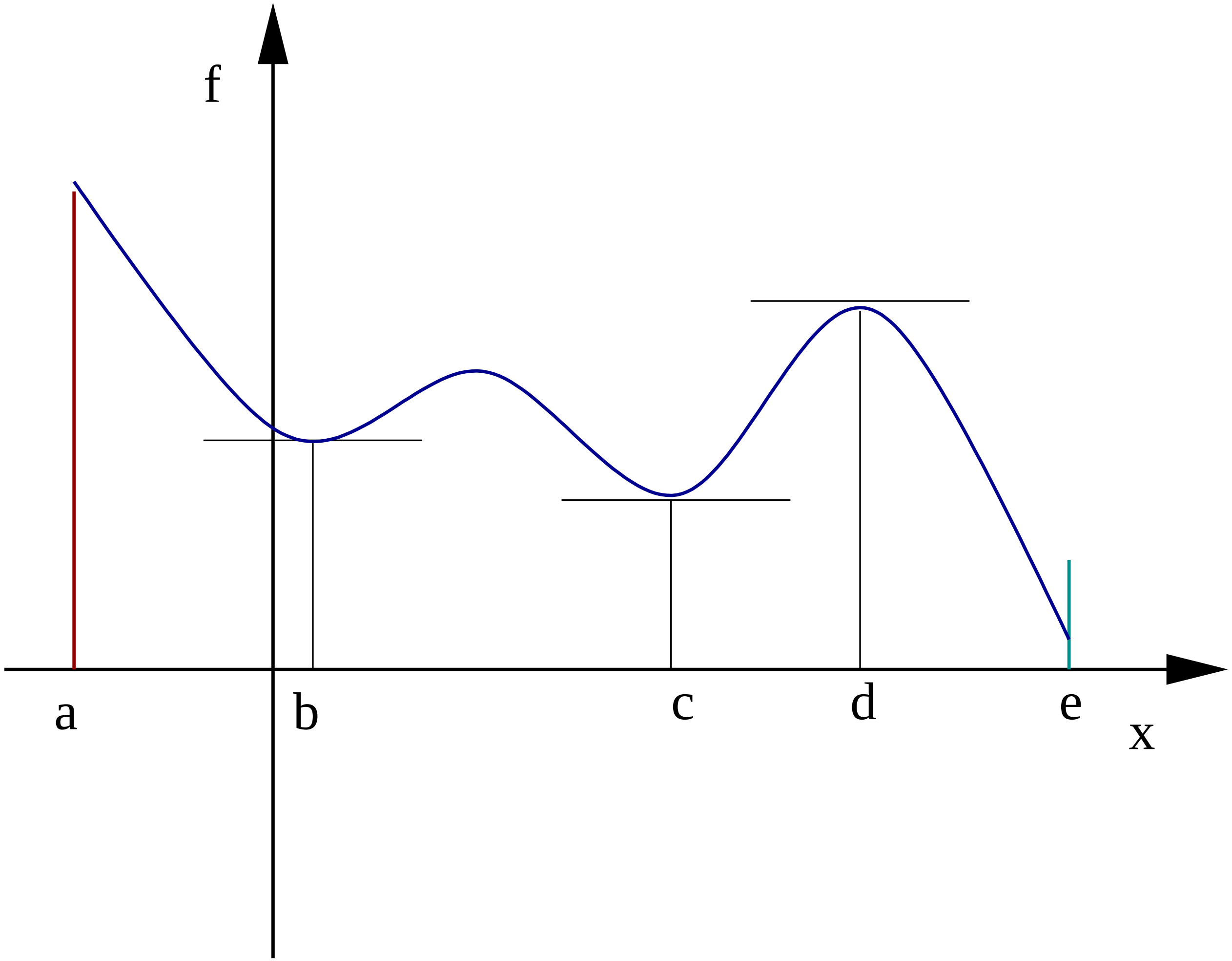

Figure 1: Global and Local Minimizers of One-Variable Function in Interval $[a \quad e]$

A differentiable function f of one variable defined on an interval $F = [a\ e]$. If an interior-point $\bar{x}$ is a local/global minimizer, then $f'(\bar{x}) = 0$; if the left-end-point $\bar{x} = a$ is a local minimizer, then $f'(a) \geq 0$; if the right-end-point $\bar{x} = e$ is a local minimizer, then $f'(e) \leq 0$. first-order necessary condition (FONC) summarizes the three cases by complementarity conditions:

$$a \leq x \leq e,\ f'(x) = y^a + y^e,\ y^a \geq 0,\ y^e \leq 0,\ y^a(x-a) = 0,\ y^b(x-e) = 0.$$

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 (SONC), which we would explored further.

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 sufficient condition (SOSC): $f''(\bar{x}) > 0$, is satisfied, then $\bar{x}$ is a local minimizer.

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

More Optimality Conditions for Unconstrained Optimization

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

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

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

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

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

The proofs would be based on 2nd-order Taylor's expansion at $\bar{\mathbf{x}}$ such that if these conditions are not satisfied, then one would be find a second-order descent-direction $\mathbf{d}$ and a small constant $\bar{\alpha} > 0$ such that $f(\bar{\mathbf{x}} + \alpha\mathbf{d}) < f(\bar{\mathbf{x}}),\ \forall 0 < \alpha \leq \bar{\alpha}$.

It may still not be sufficient, e.g., $f(x) = x^3$.

General Constrained Optimization

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

We have dealt the case when the feasible region is convex polyhedron.

We now study the case that the only assumption is that all functions are in C^1, and C^2 later, either convex or noconvex.

We again establish optimality conditions to qualify/verify any local optimizers. These conditions give us qualitative structures of (local) optimizers and lead to quantitative algorithms to find a numerical optimizer.

Lagrangian Function of Constrained Optimization

It is convenience to introduce the Lagrangian Function associated with constrained optimization:

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

where multipliers $\mathbf{y}$ is "free" and $\mathbf{s} \geq \mathbf{0}$.

Lagrangian Function can be viewed as a function aggregated the original objective function with the penalized constraint functions.

In theory, one can adjust the penalty multipliers $(\mathbf{y}, \mathbf{s} \geq \mathbf{0})$ to repeatedly solve the following so-called Lagrangian Relaxation Problem:

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

Hypersurface and Implicit Function Theorem

Consider the (intersection) of Hypersurfaces (vs. Hyperplanes):

$$\{\mathbf{x} \in R^n : \mathbf{h}(\mathbf{x}) = \mathbf{0} \in R^m, \; m \leq n\}$$

When functions $h_i(\mathbf{x})$s are C^1 functions, we say the surface is smooth.

For a point $\bar{\mathbf{x}}$ on the surface, we call it a regular point if $\nabla \mathbf{h}(\bar{\mathbf{x}})$ have rank m or the rows are linearly independent. For example, $(0; \; 0)$ is not a regular point of $\{(x_1; \; x_2) \in R^2 : \; x_1^2 + (x_2 - 1)^2 - 1 = 0, \; x_1^2 + (x_2 + 1)^2 - 1 = 0\}$.

Based on the Implicit Function Theorem, if $\bar{\mathbf{x}}$ is a regular point and $m < n$, then for every $\mathbf{d} \in \mathcal{T}_{\bar{\mathbf{x}}} = \{\mathbf{z} : \; \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{z} = \mathbf{0}\}$ there exists a curve $\mathbf{x}(t)$ on the hypersurface, parametrized by a scalar t in a sufficiently small interval $[-a \quad a]$, such that

$$\mathbf{h}(\mathbf{x}(t)) = \mathbf{0}, \quad \mathbf{x}(0) = \bar{\mathbf{x}}, \quad \dot{\mathbf{x}}(0) = \mathbf{d}.$$

$\mathcal{T}_{\bar{\mathbf{x}}}$ is called the tangent linear sub-space of the constraints at $\bar{\mathbf{x}}$.

$$\min \quad (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad (x_1)^2/4 + (x_2)^2 - 1 = 0$$

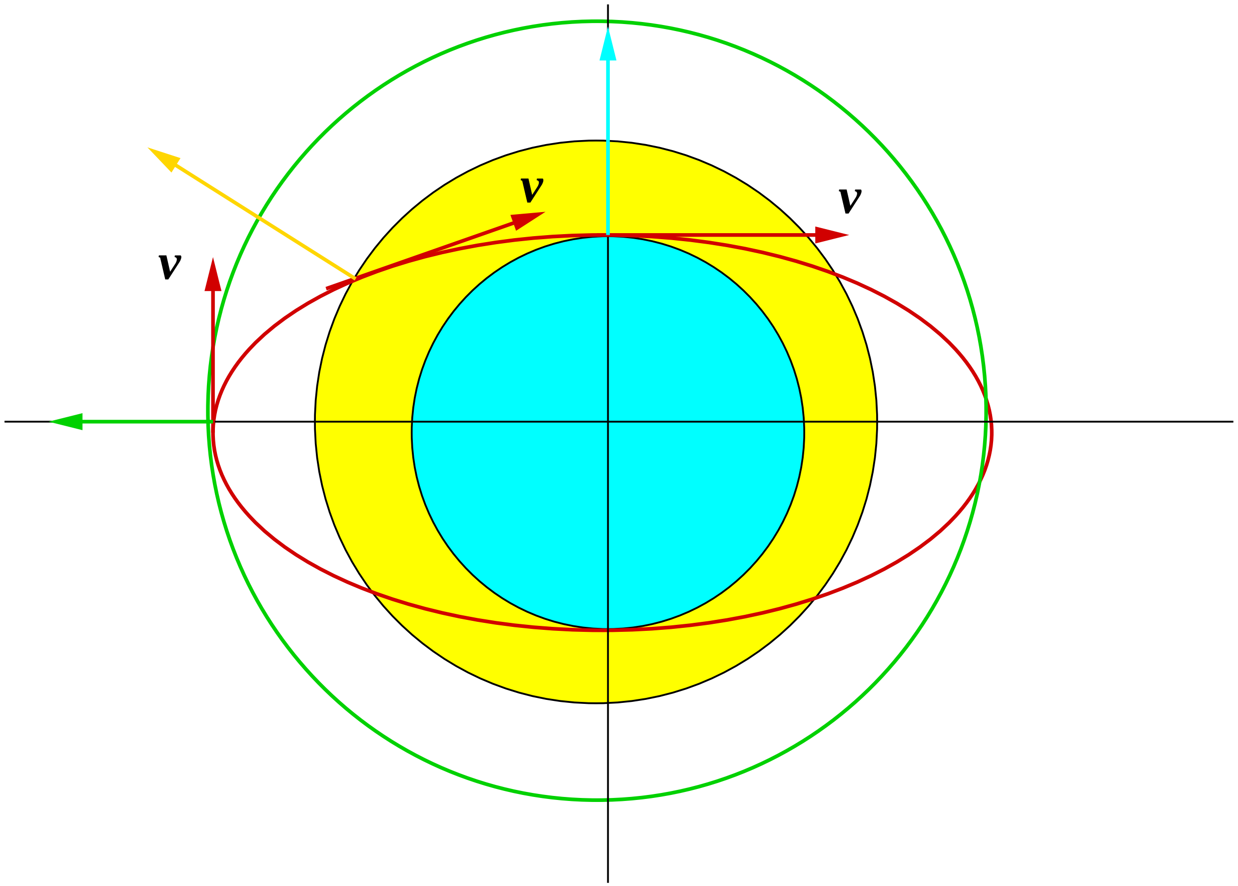

Figure 2: A Nonlinear Equality Constrained Minimization with Constraint Tangents

First-Order Necessary Conditions for Constrained Optimization I

Lemma 1 *Let $\bar{\mathbf{x}}$ be a feasible solution and a regular point of the hypersurface of*

$$\{\mathbf{x}: \ \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ c_i(\mathbf{x}) = 0, i \in \mathcal{A}_{\bar{\mathbf{x}}}\}$$

where active-constraint set $\mathcal{A}_{\bar{\mathbf{x}}} = \{i : \ c_i(\bar{\mathbf{x}}) = 0\}$. If $\bar{\mathbf{x}}$ is a (local) minimizer of (GCO), then there must be no $\mathbf{d}$ to satisfy linear constraints:

$$\begin{aligned} \nabla f(\bar{\mathbf{x}})\mathbf{d} &< 0 \\ \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{d} &= \mathbf{0} \in R^m, \\ \nabla c_i(\bar{\mathbf{x}})\mathbf{d} &\geq 0, \ \forall i \in \mathcal{A}_{\bar{\mathbf{x}}}. \end{aligned}$$

This lemma is trivial when constraints are linear since $\mathbf{d}$ is a feasible direction, but needs more work otherwise since there is no feasible direction when constraints are nonlinear.

Proof

Suppose we have a $\bar{\mathbf{d}}$ satisfies all linear constraints. Then $\nabla f(\bar{\mathbf{x}})\bar{\mathbf{d}} < 0$ so that $\bar{\mathbf{d}}$ is a descent-direction vector. Denote the active-constraint set at $\bar{\mathbf{d}}$ among the linear inequalities by $\mathcal{A}^d_{\bar{\mathbf{x}}}$ $(\subset \mathcal{A}_{\bar{\mathbf{x}}})$. Then, $\bar{\mathbf{x}}$ remains a regular point of hypersurface of

$$\{\mathbf{x} : \ \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ c_i(\mathbf{x}) = 0, i \in \mathcal{A}^d_{\bar{\mathbf{x}}}\}.$$

Thus, there is a curve $\mathbf{x}(t)$ such that

$$\mathbf{h}(\mathbf{x}(t)) = \mathbf{0}, \quad c_i(\mathbf{x}(t)) = 0, \ i \in \mathcal{A}^d_{\bar{\mathbf{x}}}, \quad \mathbf{x}(0) = \bar{\mathbf{x}}, \quad \dot{\mathbf{x}}(0) = \bar{\mathbf{d}},$$

for $t \in [-a \quad a]$ of a sufficiently small positive constant a.

Also, $\nabla c_i(\bar{\mathbf{x}})\bar{\mathbf{d}} > 0, \ \forall i \not\in \mathcal{A}^d_{\bar{\mathbf{x}}}$ and $c_i(\bar{\mathbf{x}}) > 0, \ \forall i \not\in \mathcal{A}_{\bar{\mathbf{x}}}$. From Taylor's theorem, $c_i(\mathbf{x}(t)) > 0$ for all $i \not\in \mathcal{A}^d_{\bar{\mathbf{x}}}$ so that $\mathbf{x}(t)$ is a feasible curve to the original (GCO) problem for $t \in [-a \quad a]$. Thus, $\bar{\mathbf{x}}$ must be also a local minimizer among all local solutions on the curve $\mathbf{x}(t)$.

Let $\phi(t) = f(\mathbf{x}(t))$. Then, $t = 0$ must be a local minimizer of $\phi(t)$ for $-a \le t \le a$ so that

$$0 = \phi'(0) = \nabla f(\mathbf{x}(0))\dot{\mathbf{x}}(0) = \nabla f(\bar{\mathbf{x}})\bar{\mathbf{d}} < 0, \ \Rightarrow \text{a contradiction.}$$

First-Order Necessary Conditions for Constrained Optimization II

Theorem 3 *(First-Order or KKT Optimality Condition) Let* $\bar{\mathbf{x}}$ *be a (local) minimizer of (GCO) and it is a regular point of* $\{\mathbf{x} : \ \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ c_i(\mathbf{x}) = 0, i \in \mathcal{A}_{\bar{\mathbf{x}}}\}$*. Then, for some multipliers* $(\bar{\mathbf{y}}, \bar{\mathbf{s}} \geq \mathbf{0})$

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

that is,

$$\nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) = \mathbf{0};$$

and (complementarity)

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

The proof is again based on the alternative system theory. The complementarity condition is from that $c_i(\bar{\mathbf{x}}) = 0$ for all $i \in \mathcal{A}_{\bar{\mathbf{x}}}$, and for $i \notin \mathcal{A}_{\bar{\mathbf{x}}}$, we simply set $s_i^* = 0$.

Second-Order Necessary Conditions for Constrained Optimization

Now in addition we assume all functions are in C^2, that is, twice continuously differentiable. Recall the tangent linear sub-space at $\bar{\mathbf{x}}$:

$$T_{\bar{\mathbf{x}}} := \{\mathbf{z} : \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{z} = \mathbf{0},\ \nabla c_i(\bar{\mathbf{x}})\mathbf{z} = 0\ \forall i \in \mathcal{A}_{\bar{\mathbf{x}}}\}.$$

Theorem 4 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (GCO) and a regular point of hypersurface $\{\mathbf{x} :\ \mathbf{h}(\mathbf{x}) = \mathbf{0},\ c_i(\mathbf{x}) = 0, i \in \mathcal{A}_{\bar{\mathbf{x}}}\}$, and let $\bar{\mathbf{y}}, \bar{\mathbf{s}}$ denote Lagrange multipliers such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ satisfies the (first-order) KKT conditions of (GCO). Then, it is necessary to have*

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

The Hessian of the Lagrangian function need to be positive semidefinite on the tangent space.

Proof

Consider again the objective function $\phi(t) = f(\mathbf{x}(t))$ on the feasible curve $\mathbf{x}(t)$. Since 0 is a (local) minimizer of $\phi(t)$,

$$0 \le \phi''(t)|_{t=0} = \dot{\mathbf{x}}(0)^T \nabla^2 f(\bar{\mathbf{x}}) \dot{\mathbf{x}}(0) + \nabla f(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0) = \mathbf{d}^T \nabla^2 f(\bar{\mathbf{x}}) \mathbf{d} + \nabla f(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0).$$

Let all active constraints (including the equality ones) be $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and differentiating equations $\bar{y}^T \mathbf{h}(\mathbf{x}(t)) = \sum_i \bar{y}_i h_i(\mathbf{x}(t)) = 0$ twice, we obtain

$$0 = \dot{\mathbf{x}}(0)^T [\sum_i \bar{y}_i \nabla^2 h_i(\bar{\mathbf{x}})] \dot{\mathbf{x}}(0) + \bar{y}^T \nabla \mathbf{h}(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0) = \mathbf{d}^T [\sum_i \bar{y}_i \nabla^2 h_i(\bar{\mathbf{x}})] \mathbf{d} + \bar{y}^T \nabla \mathbf{h}(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0).$$

Let the second expression subtracted from the first one on both sides and use the FONC:

$$\begin{aligned} 0 &\le \mathbf{d}^T \nabla^2 f(\bar{\mathbf{x}}) \mathbf{d} - \mathbf{d}^T [\textstyle\sum_i \bar{y}_i \nabla^2 h_i(\bar{\mathbf{x}})] \mathbf{d} + \nabla f(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0) - \bar{y}^T \nabla \mathbf{h}(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0) \\ &= \mathbf{d}^T \nabla^2 f(\bar{\mathbf{x}}) \mathbf{d} - \mathbf{d}^T [\textstyle\sum_i \bar{y}_i \nabla^2 h_i(\bar{\mathbf{x}})] \mathbf{d} \\ &= \mathbf{d}^T \nabla^2_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) \mathbf{d}. \end{aligned}$$

Note that this inequality holds for every $\mathbf{d} \in T_{\bar{\mathbf{x}}}$.

Second-Order Sufficient Conditions for GCO

Theorem 5 *Let $\bar{\mathbf{x}}$ be a regular point of (GCO) and let $\bar{\mathbf{y}}, \bar{\mathbf{s}}$ be the Lagrange multipliers such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ satisfies the (first-order) KKT conditions of (GCO). Then, if in addition*

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

then $\bar{\mathbf{x}}$ is a local minimizer of (GCO).

See the proof in Chapter 11.8 of LY.

$$\min \quad (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad -(x_1)^2/4 - (x_2)^2 + 1 \le 0$$

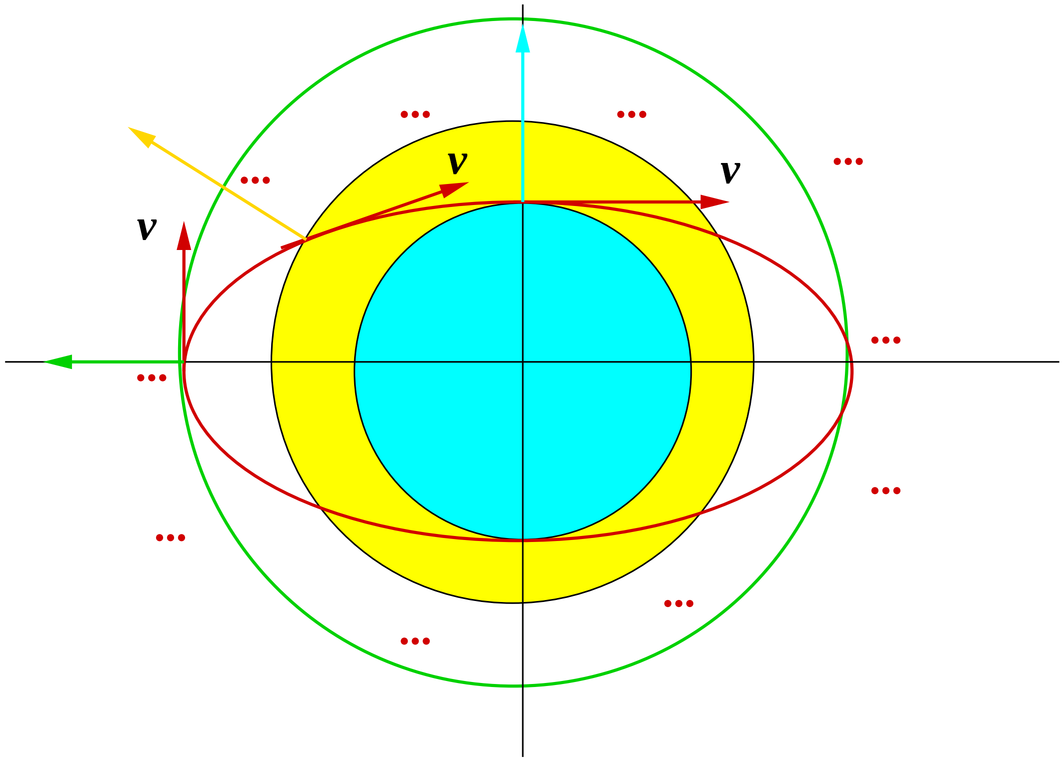

Figure 3: FONC and SONC for Constrained Minimization

$$L(x_1, x_2, y) = (x_1)^2 + (x_2)^2 - y(-(x_1)^2/4 - (x_2)^2 + 1),$$

$$\nabla_x L(x_1, x_2, y) = (2x_1(1 + y/4),\ 2x_2(1 + y)),$$

$$\nabla_x^2 L(x_1, x_2, y) = \begin{pmatrix} 2(1 + y/4) & 0 \\ 0 & 2(1 + y) \end{pmatrix}$$

$$T_{\mathbf{x}} := \{(z_1,\ z_2) :\ (x_1/4)z_1 + x_2 z_2 = 0\}.$$

We see that there are two possible values for y: either -4 or -1, which lead to total four KKT points:

$$\begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -4 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ -4 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}.$$

Consider the first KKT point:

$$\nabla_x^2 L(2,0,-4) = \begin{pmatrix} 0 & 0 \\ 0 & -6 \end{pmatrix}, \; T_{\bar{\mathbf{x}}} = \{(z_1,\ z_2) :\ z_1 = 0\}$$

Then the Hessian is not positive semidefinite on $T_{\bar{\mathbf{x}}}$ since

$$\mathbf{d}^T \nabla_x^2 L(2,0,-4)\mathbf{d} = -6d_2^2 \le 0.$$

Consider the third KKT point:

$$\nabla_x^2 L(0,1,-1) = \begin{pmatrix} 3/2 & 0 \\ 0 & 0 \end{pmatrix}, \; T_{\bar{\mathbf{x}}} = \{(z_1,\ z_2) :\ z_2 = 0\}$$

Then the Hessian is positive definite on $T_{\bar{\mathbf{x}}}$ since

$$\mathbf{d}^T \nabla_x^2 L(0,0,-1)\mathbf{d} = (3/2)d_1^2 > 0,\ \forall \mathbf{0} \neq \mathbf{d} \in T_{\bar{\mathbf{x}}}.$$

Summary Theorem of KKT Conditions for GCO

We now consider optimality conditions for problems having three types of inequalities:

$$\text{(GCO)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i\mathbf{x}) \quad (\leq, =, \geq) \quad 0, \; i = 1, ..., m, \end{array}$$

For any feasible point $\mathbf{x}$ of (GCO) define the active constraint set by $\mathcal{A}_{\mathbf{x}} = \{i : \; c_i(\mathbf{x}) = 0\}$.

Let $\bar{\mathbf{x}}$ be a local minimizer for (GCO) and $\bar{\mathbf{x}}$ is a regular point on the hypersurface of the active constraints

Then there exist multipliers $\bar{\mathbf{y}}, \bar{\mathbf{z}}$ such that

$$\begin{array}{rcl} \nabla f(\bar{\mathbf{x}}) & = & \bar{\mathbf{y}}^T \nabla \mathbf{c}(\bar{\mathbf{x}}) \\ \bar{y}_i & (\leq, '\text{free}', \geq) & 0, \; i = 1, ..., m, \\ \bar{y}_i c_i(\bar{\mathbf{x}}) & = & 0. \end{array}$$

Test Positive Semidefiniteness in a Subspace I

In the second-order test, we typically like to know whether or not

$$\mathbf{d}^T Q \mathbf{d} \geq 0, \ \forall \mathbf{d}, \text{ s.t. } A\mathbf{d} = \mathbf{0}$$

for a given symmetric matrix Q and a rectangle matrix A. (In this case, the subspace is the null space of matrix A.) This test itself might be a nonconvex optimization problem.

But it is known that $\mathbf{d}$ is in the null space of matrix A if and only if

$$\mathbf{d} = (I - A^T(AA^T)^{-1}A)\mathbf{u} = P_A\mathbf{u}$$

for some vector $\mathbf{u} \in R^n$, where P_A is called the projection matrix of A. Thus, the test becomes whether or not

$$\mathbf{u}^T P_A Q P_A \mathbf{u} \geq 0, \ \forall \mathbf{u} \in R^n,$$

that is, we just need to test positive semidefiniteness of $P_A Q P_A$ as usual.

Test Positive Semidefiniteness in a Subspace II

Another way is to apply SDP relaxation:

$$\begin{array}{lll} (SDP) & \min & Q \bullet D \\ & \text{s.t.} & A_i^T A_i \bullet D = 0; \forall i \\ & & D \succeq \mathbf{0}, \end{array}$$

where A_i is the ith row vector of A. The objective value is bounded below by 0 if the dual has a feasible solution:

$$\begin{array}{lll} (SDD) & \min & \mathbf{0}^T \mathbf{y} \\ & \text{s.t.} & Q - \sum_i y_i A_i^T A_i \succeq \mathbf{0}. \end{array}$$