Optimality Conditions for Linear and Nonlinear Optimization via the Lagrange Function

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye
Chapters 7.1-7.2, 11.1-11.7

Mathematical Optimization Problems

Recall Mathematical Optimization Problem Form:

$$(\text{MOP}) \quad \begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in F. \end{array}$$

The first question: does the problem have a feasible solution, that is, a solution that satisfies all the constraints of the problem, that is, in F .

The second question: How does one recognize or certify a (local) optimal solution? We answered it for LP by developing Optimality Conditions from the LP duality and Complementarity.

But what about a generally nonlinearly constrained and objective optimization problem? We need more general Optimality Condition Theory.

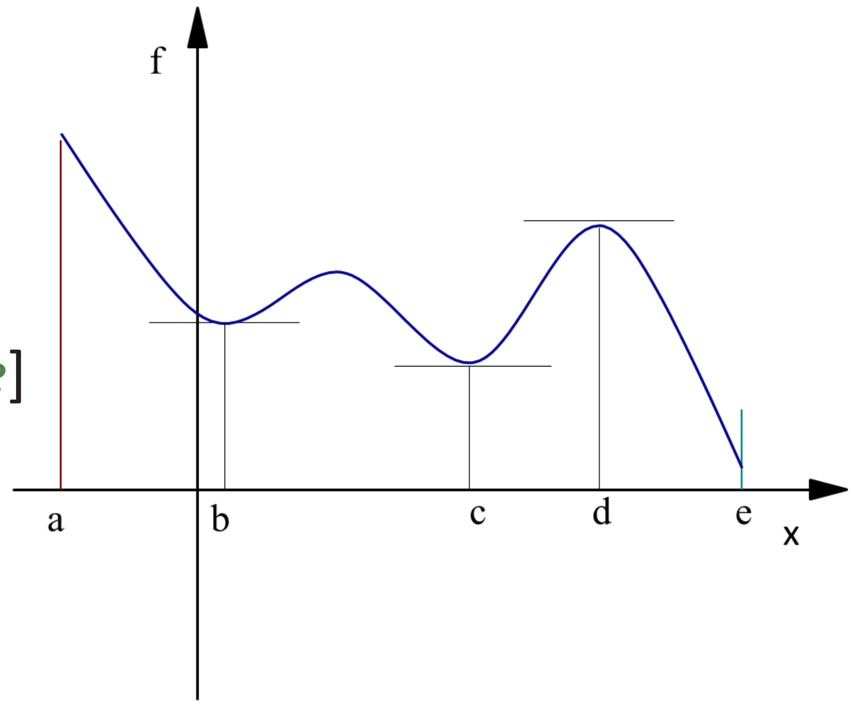
Remarks of Optimality Condition Theory

- 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 (or KKT) conditions and second-order optimality conditions. The solution **x**, together with the multipliers **y**, is called an KKT solution/point if they satisfy the KKT conditions.
- For convex optimization (CO), first-order or KKT optimality conditions suffice (under mild technical assumptions). Also, these set of conditions are necessary for nonlinearly constrained optimization under some mild technical assumptions.

Consider Minimization Problems with One Variable

Mathematical Optimization Problem Form:

$$(\text{MOP1}) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & a \le x \le e. \end{array}$$



Which point in the constraint interval [*a* *e*] can be possibly a minimal solution of *f* .

Test of *a*: *f'(a) ≥ 0 (feasible direction is +)*
Test of *e*: f'(e) ≤ 0 *(feasible direction is -)*
Test of any point strictly inside the interval: *f'(x)=0 (feas. direction is +/-)*

To summarize the three cases, one can introduce two Lagrange multipliers $y_a \ge 0$ and $y_e \ge 0$ so that the optimality conditions can be characterized as:

$$f'(x)-y_a+y_e=0,\ y_a(x-a)=0,\ y_e(e-x)=0,$$

This is called the (first-order) Optimality/KKT Condition of the problem.

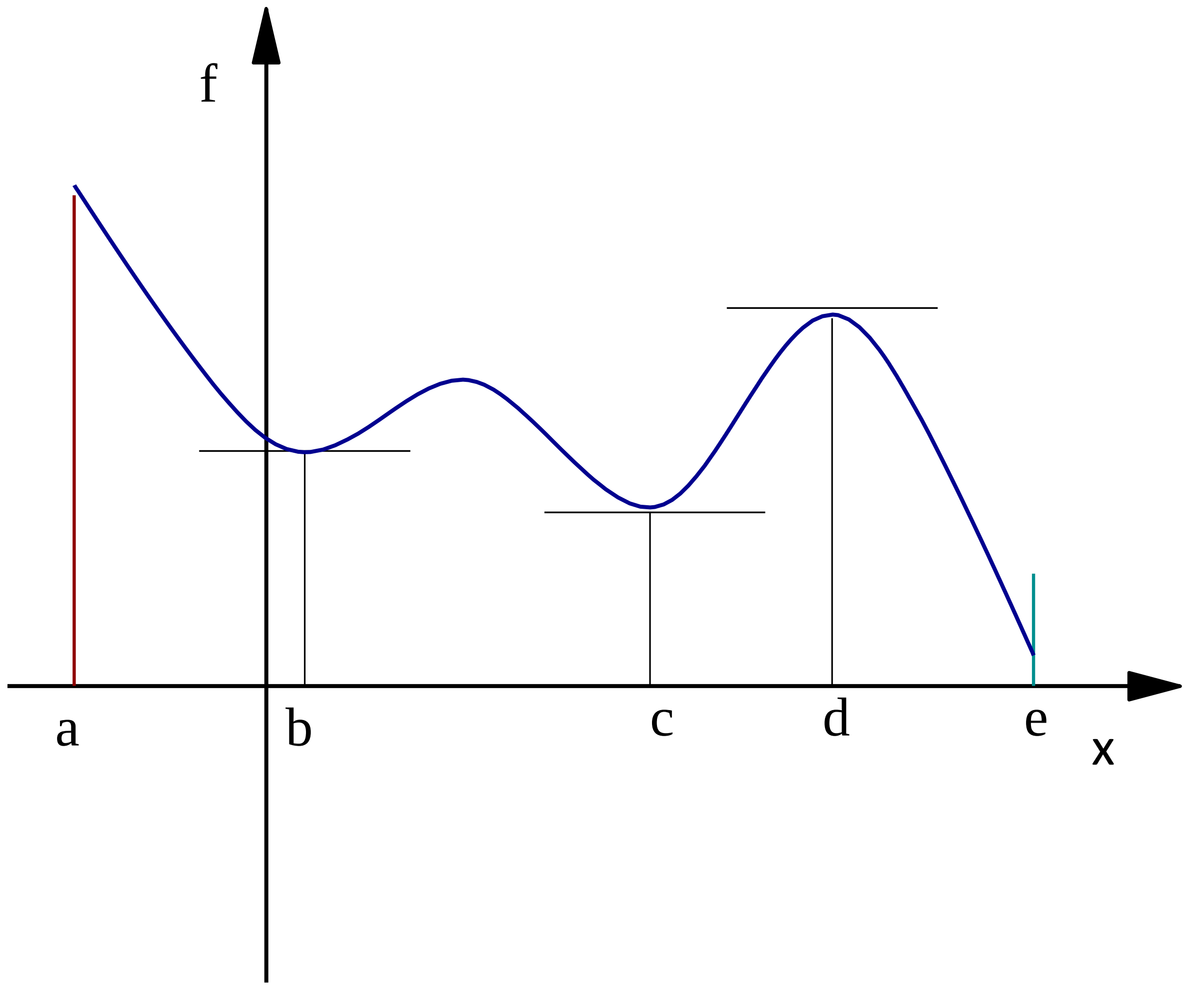


Figure : Possible local minimizers of one-variable function or KKT points/solutions: b), c), d) and e)

Optimality Conditions via the Lagrange Function

Mathematical Optimization Problem Form:

$$\text{(MOP1)} \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & x - a \ge 0 \;\; (y_a \ge 0),\; e - x \ge 0 \;\; (y_e \ge 0). \end{array}$$

The Lagrange function or Lagrangian:

$$L(x, y_a \ge 0, y_e \ge 0) = f(x) - y_a(x - a) - y_e(e - x)$$

Therefore, together with complementarity, the first equation of the optimality conditions can be simply written as:

$$L'_x(x, y_a \ge 0, y_e \ge 0) = f'(x) - y_a + y_e = 0,\; y_a(x - a) = 0,\; y_e(e - x) = 0$$

Consider a specific function case

$$L(x, y_a \ge 0, y_e \ge 0) = x^2 - y_a(x - a) - y_e(e - x)$$

$[a \;\; e]=[-2 \;\; -1]$: $x=-1$, $y_a =$, $y_e =$,

$[a \;\; e]=[-1 \;\; 1]$: $x=0$, $y_a =$, $y_e =$

$[a \;\; e]=[1 \;\; 2]$: $x=1$, $y_a =$, $y_e =$ (and their physical interpretation?)

Lagrange Multipliers and Functions for Multi-Variate Linear and Nonlinear Optimization

min $\quad f(\boldsymbol{x})$

s.t. $\quad c_i(\boldsymbol{x})\ (\geq,=,\leq)\ 0,\ \text{i}=1,\dots,\text{m}$

Assign each constraint a multiplier y_i, and its sign satisfies

$$y_i\ (\geq,\text{"free"},\leq)\ 0,\ i=1,\dots,m$$

The Lagrange Function

$$L(\mathbf{x},\mathbf{y}) = \text{f}(\boldsymbol{x}) - \sum_i y_i c_i(\mathbf{x})$$

The (First-Order Necessary) Optimality Conditions

Original Decision Constraints (ODC)

$c_i(\boldsymbol{x})$ ($\geq,=,\leq$) 0, i=1,…,m

Multiplier Sign Conditions (MSC)

y_i ($\geq$,"free",$\leq$) 0, $i=1,\ldots,m$

Lagrange Derivative Conditions (LDC)

$\partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$, for all $j=1,\ldots,n$.

Complementarity Slackness Condition (CSC)

$y_i c_i(\mathbf{x}) = 0$, for each inequality constraint i.

For maximization, just flip the sign of multipliers, and every condition remains the same.

Short-Cut in dealing

ODC: $x_j \geq 0$

LDC: $\partial L(\mathbf{x},\mathbf{y})/\partial x_j \geq 0$

CSC: $x_j \partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$

Optimality Conditions for Unconstrained Problems

Consider the unconstrained problem, where f is differentiable on R^n,

$$(\text{UP}) \qquad \begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in R^n. \end{array}$$

Theorem 1 *Let* $\boldsymbol{x}$ *be a (local) minimizer of (UP) where* f *is continuously differentiable at* $\boldsymbol{x}$ *. Then*

$$\nabla f(\mathbf{x}) = \mathbf{0}.$$

These conditions are sufficient if $f(.)$ *is a convex function of* $\boldsymbol{x}$*.*

Quadratic Optimization

Quadratic Function:

$$f(x) = x^T Q x - 2c^T \mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}x_i x_j - 2\sum_{j=1}^{n} c_j x_j$$

A minimizer or maximizer **x** must satisfy

$$\nabla f(\mathbf{x}) = 2Q\mathbf{x} - 2\mathbf{c} = 0 \quad \text{or} \quad Q\,\mathbf{x} = \mathbf{c}.$$

Pricing Example with demand functions:

$$d_1(\boldsymbol{x}) = 2 - x_1 + x_2$$

$$d_2(\boldsymbol{x}) = 3 - 2x_2 + x_1$$

$$Profit\ (\boldsymbol{x}) = x_1 d_1(\boldsymbol{x}) + x_2 d_2(\boldsymbol{x}) = x_1(-x_1 + x_2) + x_2(-2x_2 + x_1) + 2x_1 + 3x_2$$

Linear Equality Constrained Problems

Consider the linear equality constrained problem, where f is differentiable on R^n,

$$\text{(LEP)} \quad \begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & A\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

$$L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b})$$

Theorem 2 *Let* $\mathbf{x}$ *be a (local) minimizer of (LEP) where function* f *is continuously differentiable at* $\mathbf{x}$ *. Then*

$$\nabla f(\mathbf{x}) - A^T \mathbf{y} = \mathbf{0}$$

for a vector $\mathbf{y} = (y_1; \ldots ; y_m) \in R^m$*, which are called Lagrange (or dual) multipliers. These conditions are sufficient if* $f(.)$ *is convex.*

The geometric interpretation: the objective gradient vector is perpendicular to (linear combination) or the objective level set tangents the constraint hyperplanes (normal directions).

Linear Equality Constrained Problems

min $\quad (x_1 - 1)^2 + (x_2 - 1)^2$

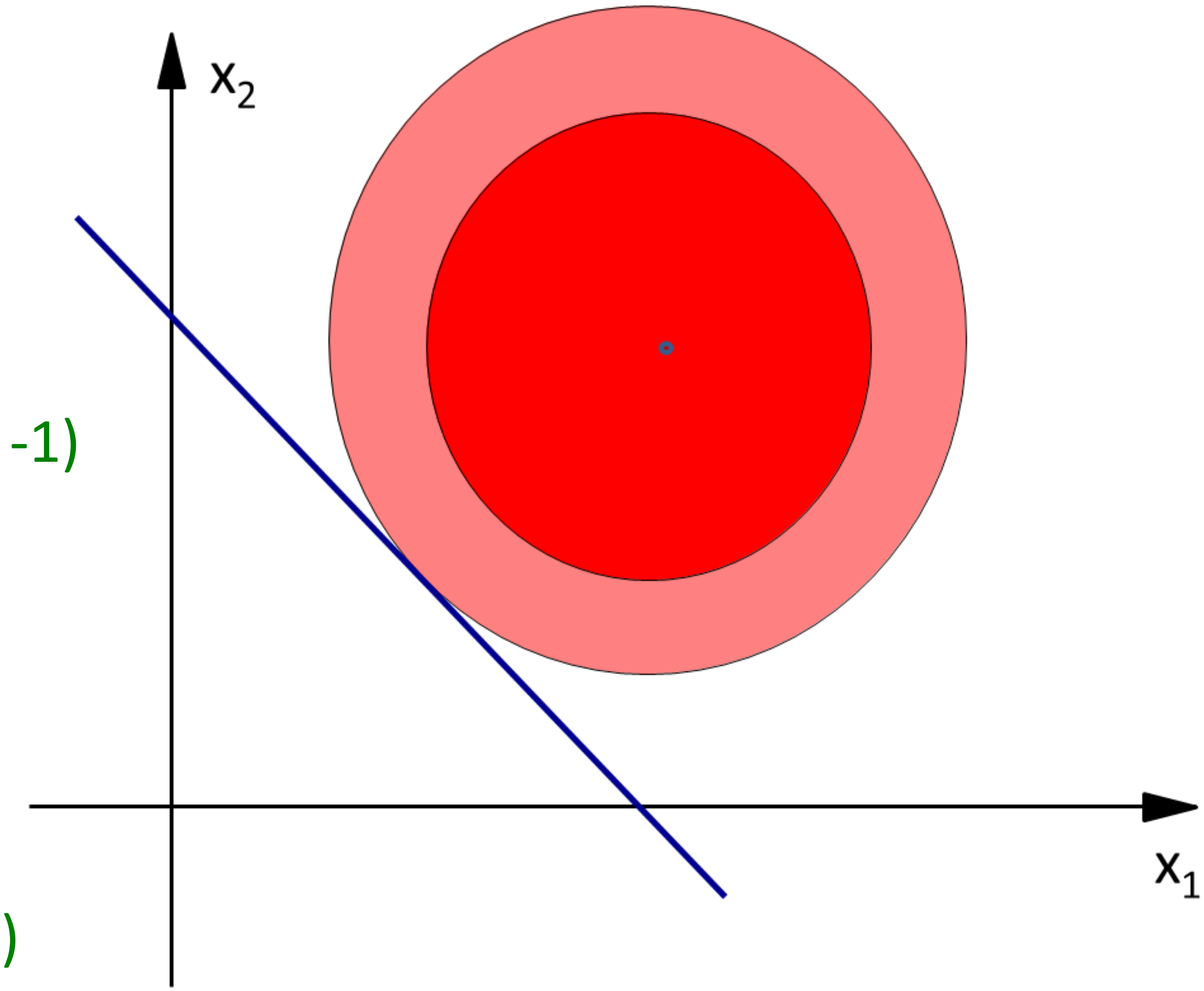
s.t. $\qquad x_1 + x_2 = 1.$

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

$2(x_1-1) = y$

$2(x_2 -1)=y,$

$(y+2)/2+(y+2)/2=1 \rightarrow$
$y=-1$ (physical interpretation?)



The objective level set tangents the constraint hyperplane

Linear Inequality Constrained Problems

Consider the linear inequality constrained problem, where f is differentiable on R^n,

$$\text{(LIP)} \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & A\mathbf{x} \geq \mathbf{b}. \end{array}$$

$$L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) \text{ where } \mathbf{y} \geq \mathbf{0}$$

Theorem 3 *Let* $\mathbf{x}$ *be a (local) minimizer of (LIP) where function* f *is continuously differentiable at* $\mathbf{x}$ *. Then*

$$\nabla f(\mathbf{x}) - A^T\mathbf{y} = \mathbf{0},\ \mathbf{y} \geq \mathbf{0}, \text{ and } y_i(A\mathbf{x} - \mathbf{b})_i = 0, \text{ for all } i=1,\ldots,m$$

for a vector $\mathbf{y} = (y_1; \ldots ; y_m) \in R^m$*, which are called Lagrange or dual multipliers. These conditions are sufficient if* $f(.)$ *is convex.*

The geometric interpretation: the objective gradient vector is a conic combination of the normal directions of the binding/active constraint hyperplanes, same as in the LP case.

Linear Inequality Constrained Example I

min $(x_1 - 1)^2 + (x_2 - 1)^2$

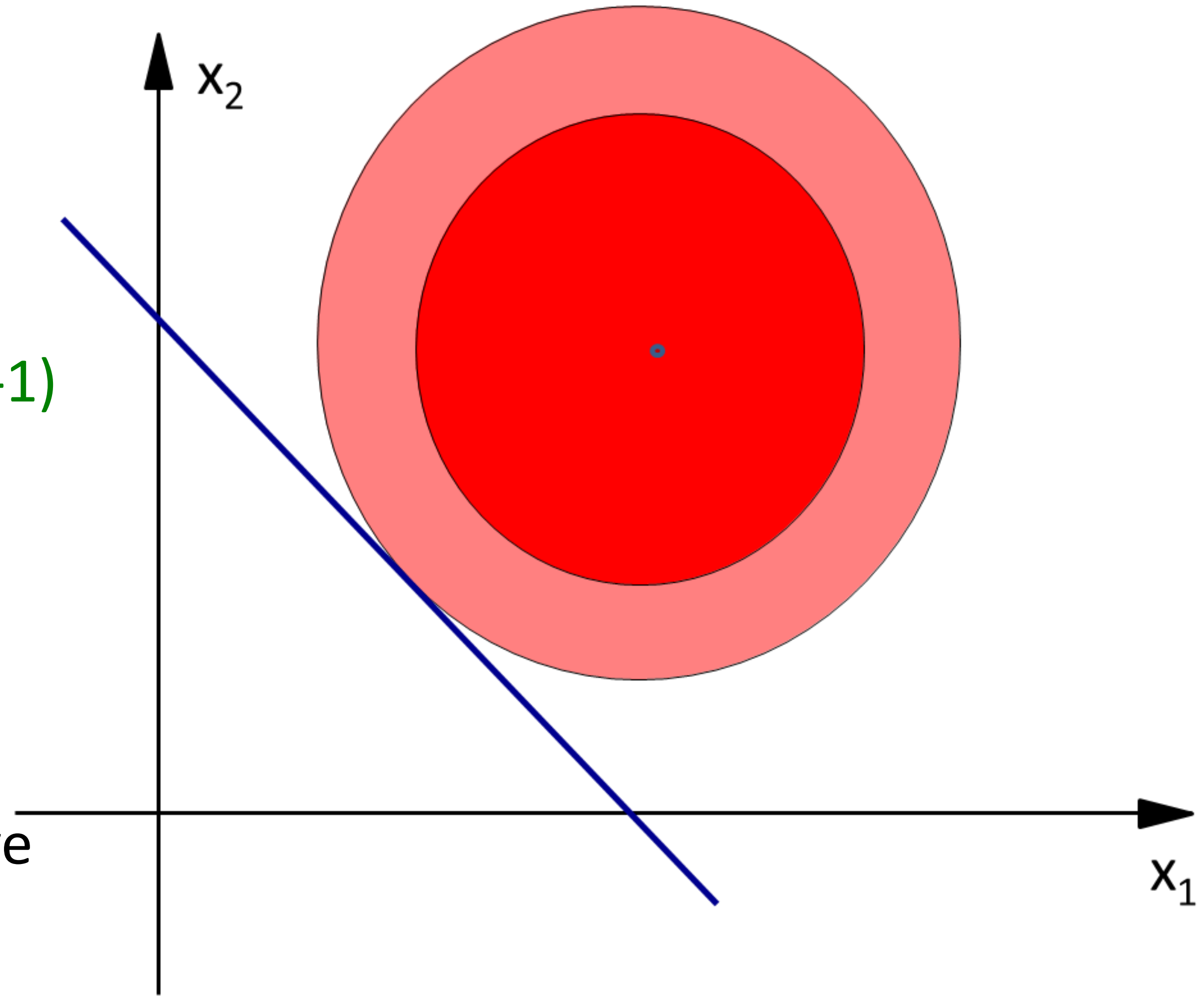
s.t. $x_1 + x_2 \geq 1.$

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

$2(x_1-1) - y = 0,$

$2(x_2 -1)-y = 0,$

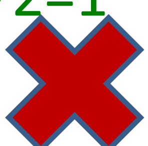
$y \geq 0,\ y(x_1+x_2 -1)=0.$



Try the constraint binding active

$(y+2)/2+(y+2)/2=1 \rightarrow$

$y=-1.$



The constraint cannot be binding or active so that we ignore it and have y=0, which leads to $x_1 = x_2 = 1$.

Linear Inequality Constrained Example II

min $\quad (x_1 - 1)^2 + (x_2 - 1)^2$

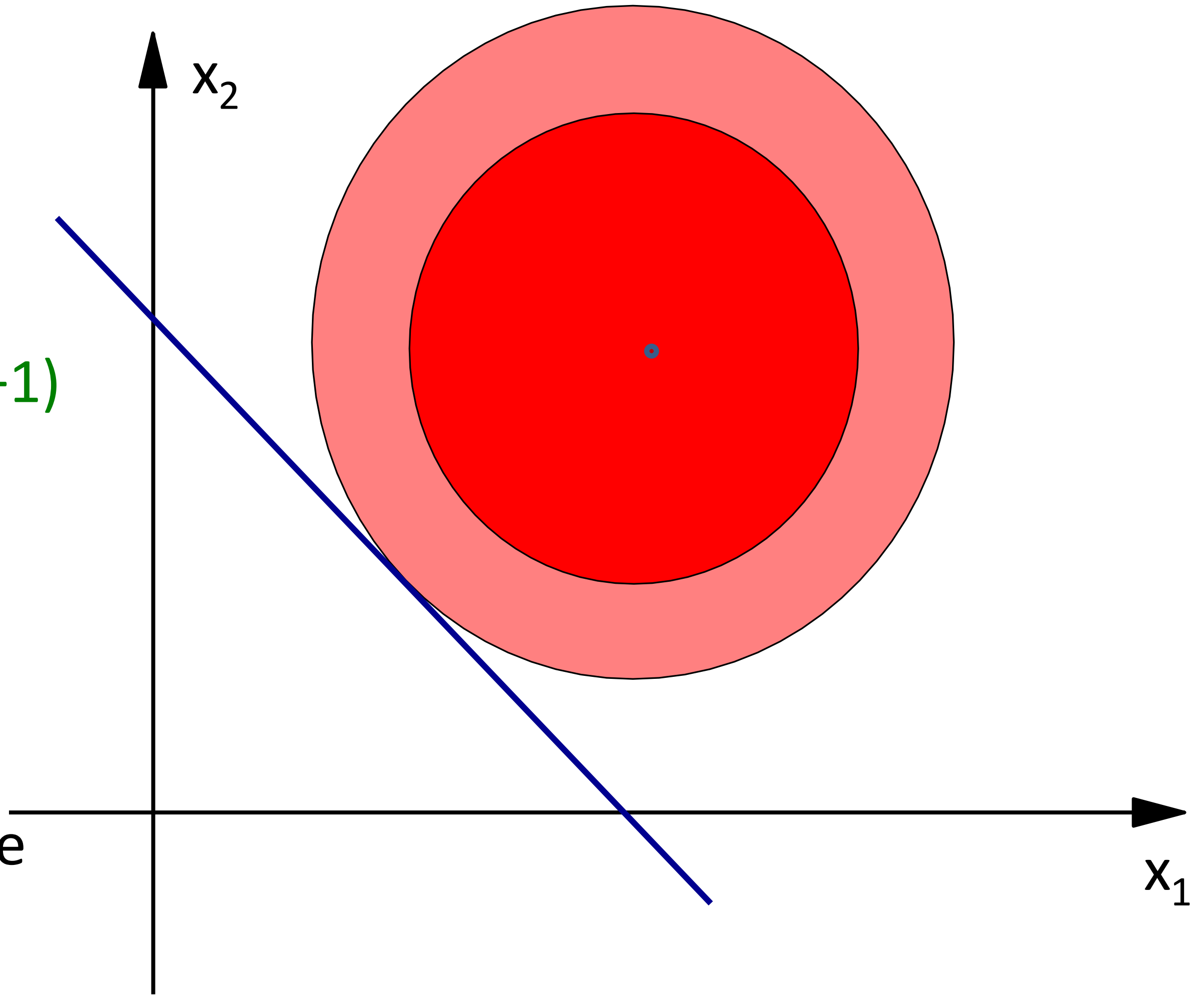
s.t. $\quad -x_1 - x_2 \geq -1.$

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

$2(x_1 - 1) - y = 0,$

$2(x_2 - 1) - y = 0,$

$y \geq 0,\ y(x_1 + x_2 - 1) = 0.$



Try the constraint binding active

$(-y+2)/2 + (-y+2)/2 = 1 \rightarrow$

$y = 1.$

The constraint will be binding or active.

Linearly Constrained Problems

Consider the linear equality and non-negativity constrained problem, where f is differentiable on R^n,

$$\text{(LENP)} \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

$$L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) - \mathbf{r}^T\mathbf{x} \text{ where } \mathbf{r} \geq \mathbf{0}$$

Theorem 4 *Let* $\mathbf{x}$ *be a (local) minimizer of (LENP) where function* f *is continuously differentiable at* $\mathbf{x}$ *. Then*

$$\nabla f(\mathbf{x}) - A^T\mathbf{y} = \mathbf{r} \geq \mathbf{0}, \quad x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j = 0, \text{ for all } j=1,..,n.$$

for a (shadow price) vector $\mathbf{y} = (y_1; \ldots; y_m) \in R^m$*, which are also called Lagrange or dual multipliers, and (reduced cost vector)* $\nabla f(\mathbf{x}) - A^T\mathbf{y}$ *. These conditions are sufficient if* $f(.)$ *is convex.*

The Log-Barrier Example

min $\quad - \log(x_1) - \log(x_2)$

s.t. $\quad x_1 + 2x_2 = 1,\ x_1, x_2 \geq 0,$

$-(1/x_1) = y$

$-(1/x_2) = 2y,$

→

$-(1/y)+2(-1/2y)=1$

$y=-2.$

→

$x_1 = 1/2$

$x_2 = 1/4.$

min $\quad - \log(x_1) - \log(x_2) - \ldots - \log(x_n)$

s.t. $\quad A\boldsymbol{x} = \boldsymbol{b},\ x_1, x_2, \ldots, x_n \geq 0,$

There is a $\boldsymbol{y}$ such that

$-(1/x_j) = \boldsymbol{a}_j^T\boldsymbol{y}$, $j=1,\ldots,n$; or

$x_j(-\boldsymbol{a}_j^T\boldsymbol{y})_j = 1$, for all j.

The optimal solution is called analytical center of the feasible region

The Constrained Quadratic Example

min $\quad -x_1(-x_1 + x_2) - x_2(-2x_2 + x_1) - 2x_1 - 3x_2$

s.t. $\quad x_1 - 2x_2 = 0,\ x_1, x_2 \geq 0,$

$2x_1 - 2x_2 - 2 = y$

$-2x_1 + 4x_2 - 3 = -2y,$

$x_1 - 2 = y$

$-3 = -2y$

$y = 1.5$

$x_1 = 3.5$

$x_2 = 1.75.$

min $\quad \boldsymbol{x}^T Q \boldsymbol{x} - 2\boldsymbol{c}^T \boldsymbol{x}$

s.t. $\quad A\boldsymbol{x} = \boldsymbol{b},\ (x_1, x_2, \ldots, x_n) \geq 0,$

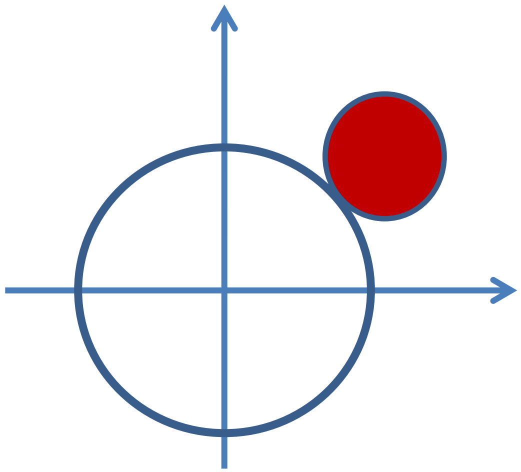
There is a $\boldsymbol{y}$ such that

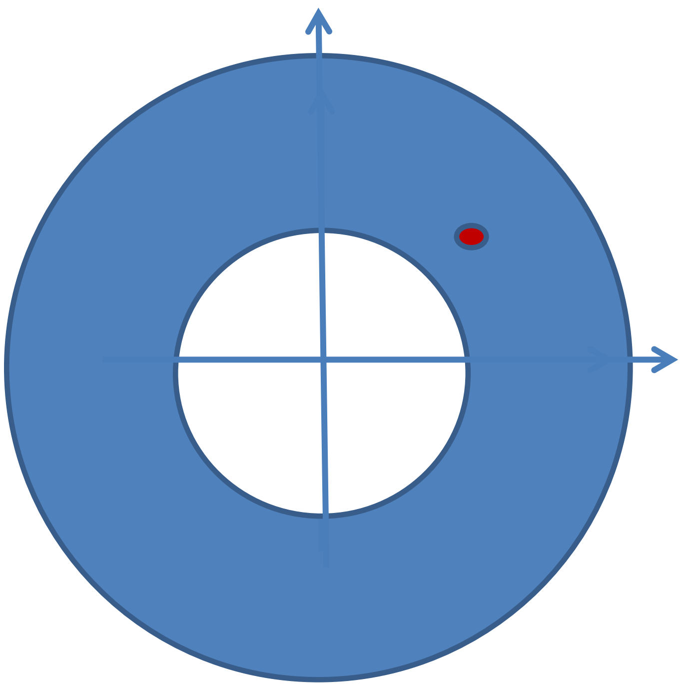
$2Q\boldsymbol{x} - 2\boldsymbol{c} \geq A^T\boldsymbol{y}$; and

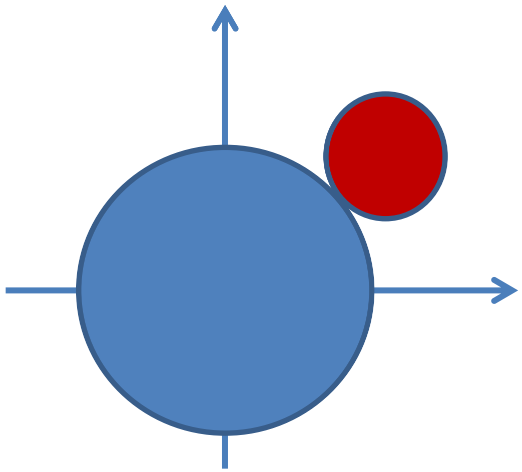
$2\boldsymbol{x}^T Q \boldsymbol{x} - 2\boldsymbol{c}^T \boldsymbol{x} = \boldsymbol{b}^T \boldsymbol{y}.$

Nonlinearly Constrained Optimization Examples

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

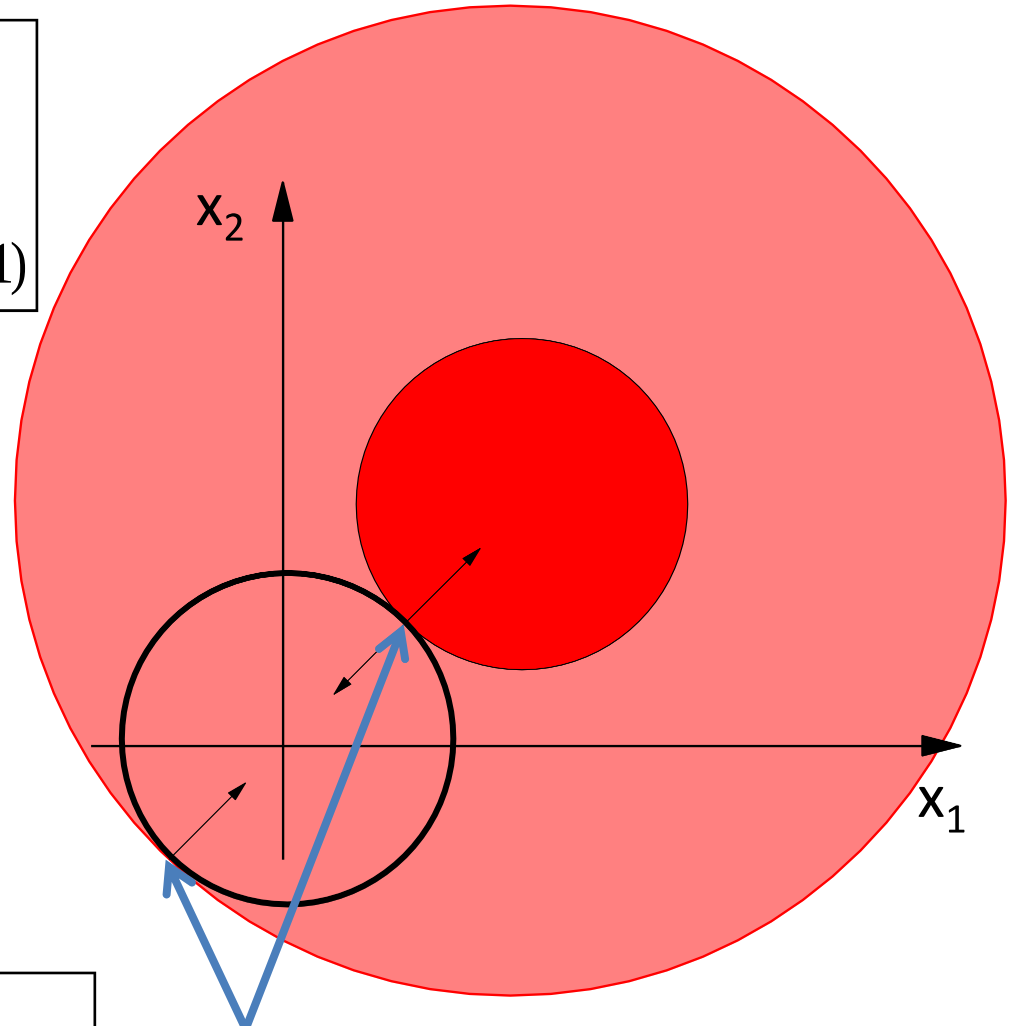






Optimality Conditions of Example I

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



KKT Conditions :

$$\begin{pmatrix} 2(x_1-1) \\ 2(x_2-1) \end{pmatrix} - \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \cdot y = 0$$

$$(x_1)^2+(x_1)^2-1=0$$

$$x_1(1-y)=1$$
$$x_2(1-y)=1$$

$$x_1 = x_2$$
$$(x_1)^2+(x_1)^2=1$$

$$x_1 = x_2 = \pm\frac{1}{\sqrt{2}}$$
$$y = 1 \mp \sqrt{2}$$

Two KKT Solutions

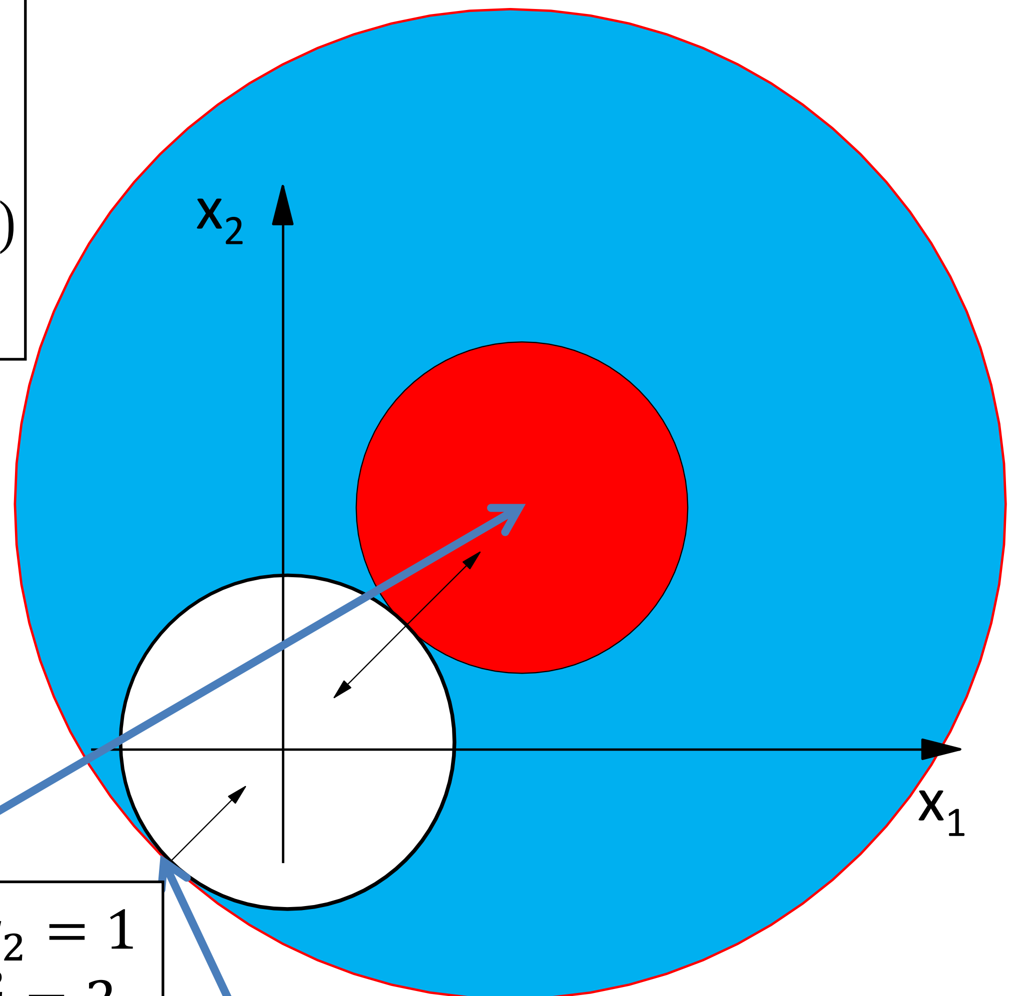
Optimality Conditions of Example II

$$\min \ (x_1-1)^2+(x_2-1)^2$$
$$\text{s.t.} \ (x_1)^2+(x_2)^2-1\geq 0$$
$$\text{LF}: (x_1-1)^2+(x_2-1)^2-y((x_1)^2+(x_2)^2-1)$$
$$(y\geq 0)$$

KKT Conditions :

$$\begin{pmatrix} 2(x_1-1) \\ 2(x_2-1) \end{pmatrix} - \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \cdot y = 0$$

$$0\leq (x_1)^2+(x_1)^2-1 \wedge y\geq 0$$



$$x_1(1-y)=1$$
$$x_2(1-y)=1$$

$y=0 \Rightarrow x_1=x_2=1$
$\& \ (x_1)^2+(x_2)^2=2$
≥ 1

$x_1 = x_2 \ \&$
$(x_1)^2+(x_2)^2=1$

$$x_1=x_2=-\frac{1}{\sqrt{2}}, y=1+\sqrt{2}$$

$$x_1=x_2=-\frac{1}{\sqrt{2}}, y=1-\sqrt{2}$$

Two KKT Sol's

Optimality Conditions of Example III

$$\min \quad (x_1-1)^2+(x_2-1)^2$$
$$\text{s.t.} \quad -(x_1)^2-(x_2)^2+1\ge 0$$
$$\text{LF}: (x_1-1)^2+(x_2-1)^2-y(1-(x_1)^2-(x_2)^2)$$
$$(y\ge 0)$$

KKT Conditions :

$$\begin{pmatrix}2(x_1-1)\\ 2(x_2-1)\end{pmatrix}+\begin{pmatrix}2x_1\\ 2x_2\end{pmatrix}\cdot y=0$$
$$0\le 1-(x_1)^2-(x_1)^2 \wedge y\ge 0$$

$$x_1(1+y)=1$$
$$x_2(1+y)=1$$

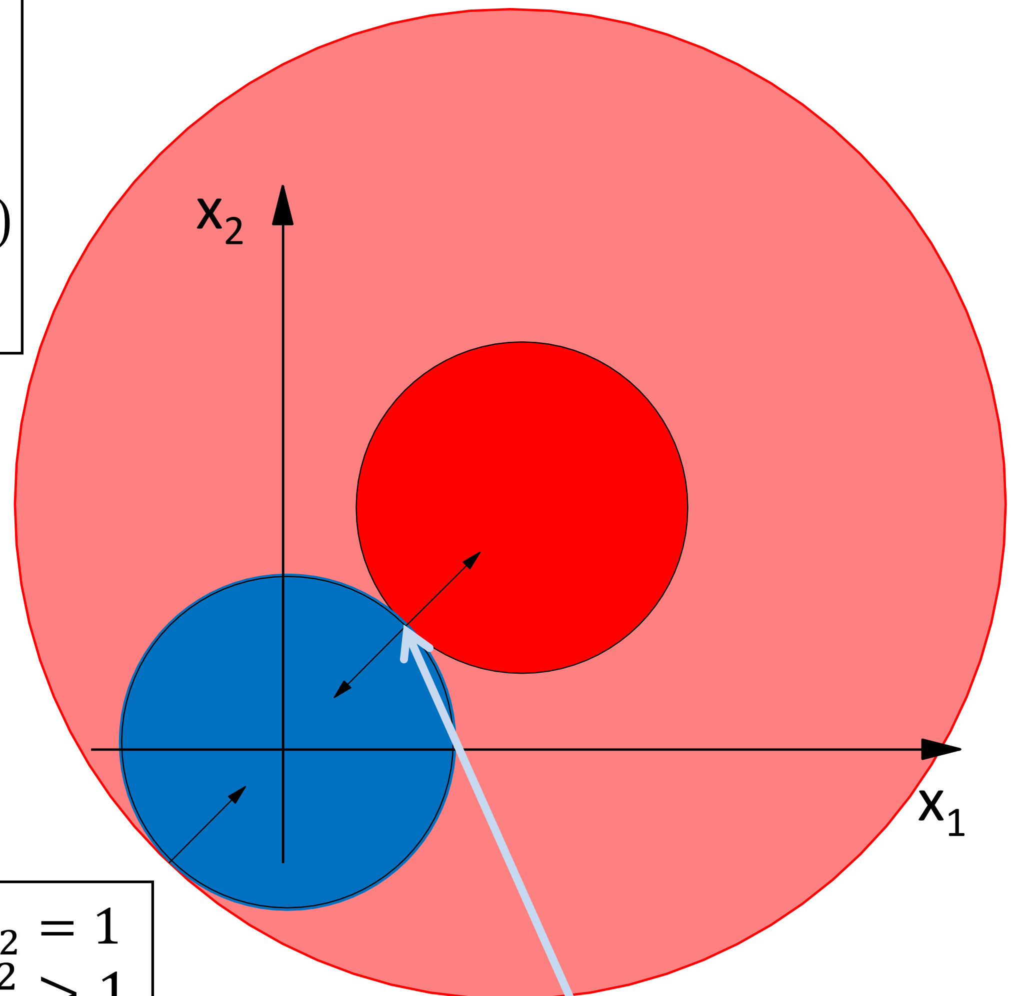
$$y=0 \Rightarrow x_1=x_2=1$$
$$\Rightarrow (x_1)^2+(x_1)^2>1$$

$$x_1=x_2$$
$$(x_1)^2+(x_1)^2=1$$

$$x_1=x_2=\frac{1}{\sqrt{2}},\ y=\sqrt{2}-1$$

$$x_1=x_2=\frac{-1}{\sqrt{2}}\ \ =-\sqrt{2}-1$$

One KKT Solution

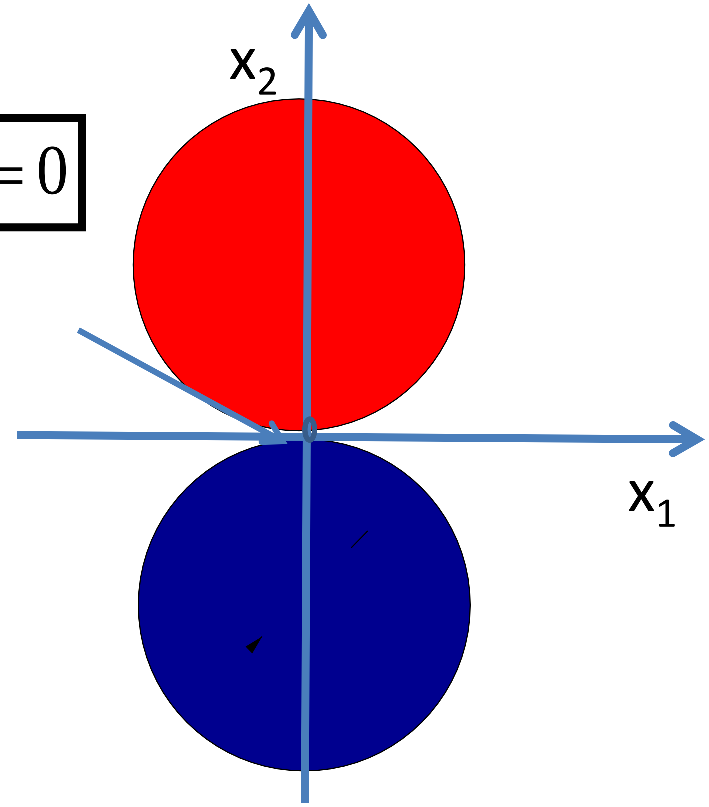


The Need Of Constraint Qualification

The KKT optimality condition or characterization of a (local) minimizer may not hold if the constraint qualification is not satisfied; that is, a minimizer may not meet the so-called "**necessary**" optimality conditions under some pathological circumstances (which occurs only when nonlinear constraints present).

$$x_1 = 0, x_2 = 0$$

$$\begin{aligned} \min \quad & x_1 \\ \text{s.t.} \quad & (x_1)^2 + (x_2 - 1)^2 - 1 \le 0, \\ & (x_1)^2 + (x_2 + 1)^2 - 1 \le 0 \end{aligned}$$



But it does not satisfy the KKT conditions

Constraint Qualification

- For an optimizer to be an KKT solution one needs some technical assumptions, called constrained and/or regularity qualifications.
- For equality constraints, the standard qualification is that the Jacobian matrix on the test solution is full rank, or the gradient vector of each equality constraint are linearly independent, at the minimizer.
- For inequality constraints, the standard qualification is that there is a feasible direction at the test solution pointing to the interior of the feasible region.
- When the problem data are randomly perturbed a little, these constraint qualifications would be met with probability one.
- These qualifications are not needed when the constraints are linear/affine.

Second Order Optimality Condition for Unconstrained Minimization

The fundamental concept of the first order necessary condition (FONC) in optimization is that there is no feasible and descent direction ***d*** at same time

The fundamental concept of the second order necessary condition (SONC) is that even the first order condition is satisfied at ***x***, then one must have also $\mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d} \geq 0$ for any feasible direction or $\nabla^2 f(\mathbf{x})$ is positive definite at ***x***.

Consider $f(x) = x^2$ and $f(x) = -x^2$, and they both satisfy the first-order necessary condition at $x = 0$, but only one of them satisfy the second-order condition.

The second order condition would be sufficient if $\mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d} > 0$ for any feasible direction **d**, while meet FONC.

Second Order Optimality Condition for Constrained Minimization

$$\min \quad f(\mathbf{x})$$

$$\text{s.t.} \quad c_i(\mathbf{x})\ (\geq,=,\leq)\ 0,\ i=1,\ldots,m$$

The Lagrange Function: $L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \sum_i y_i c_i(\mathbf{x}),\ y_i\ (\geq,\text{"free"},\leq)\ 0,\ i=1,\ldots,m$

The second order necessary condition (SONC) is that even the FONC is satisfied at $\mathbf{x}$ and $\mathbf{y}$, then they must meet also

$$\mathbf{d}^T \nabla_x^2 L(\mathbf{x},\mathbf{y})\mathbf{d} \geq 0 \quad \text{for all } \mathbf{d} \text{ such that}$$

$$\nabla c_i(\mathbf{x})\mathbf{d}=0 \quad \text{for all } i \text{ in } A(\mathbf{x})$$

where $A(\mathbf{x})$ represents the index set of all active constraints (that is, the index set of all i such that $c_i(\mathbf{x})=0$).

The second-order condition would be sufficient if $\mathbf{d}^T \nabla_x^2 L(\mathbf{x},\mathbf{y})\mathbf{d} > 0$ while meet FONC.

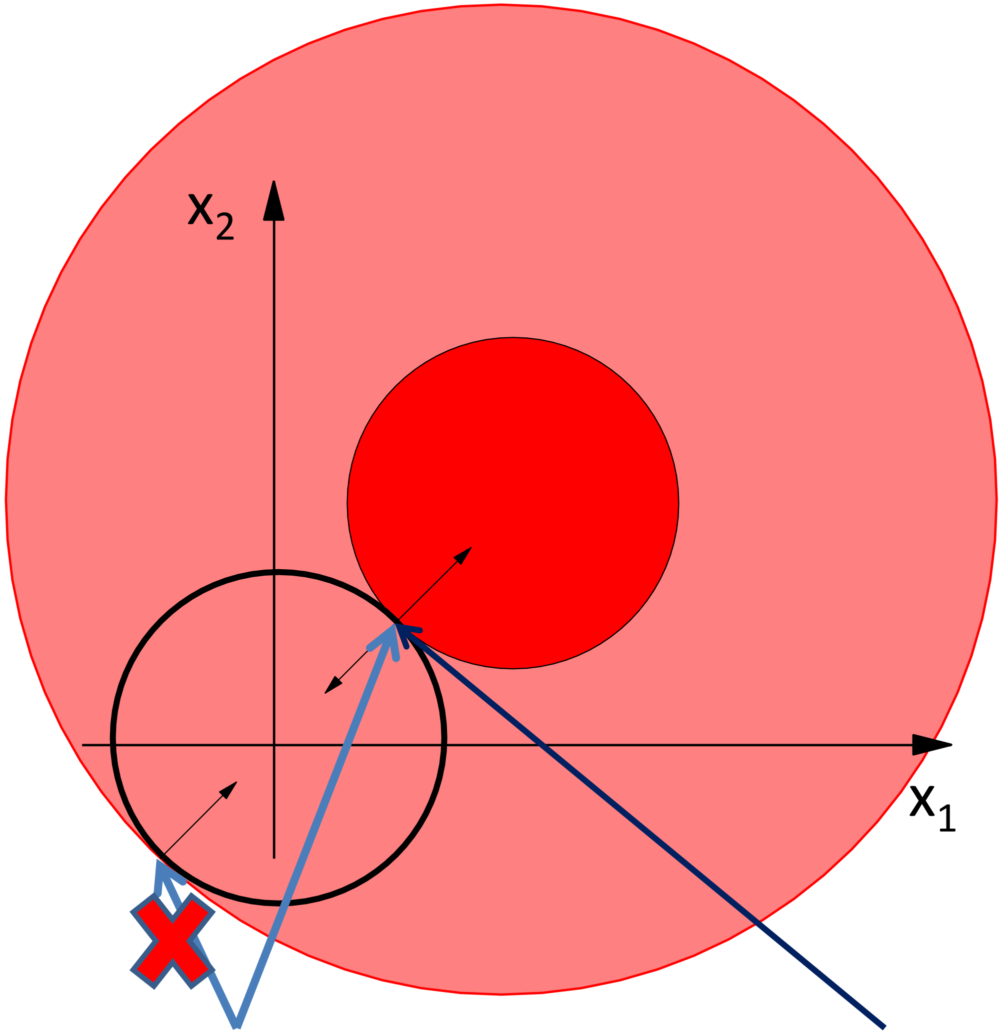
Who satisfies the second order necessary condition?

$$\begin{aligned} \min \quad & (x_1-1)^2+(x_2-1)^2 \\ \text{s.t.} \quad & (x_1)^2+(x_2)^2 \; -1=0 \end{aligned}$$

$$L(x,y)=(x_1-1)^2+(x_2-1)^2 \\ \qquad -y((x_1)^2+(x_2)^2-1)$$

$$\nabla_x L(x,y)=\begin{pmatrix} 2(x_1-1)-2yx_1 \\ 2(x_2-1)-2yx_2 \end{pmatrix}$$

$$\nabla_x^2 L(x,y)=\begin{pmatrix} 2(1-y) & 0 \\ 0 & 2(1-y) \end{pmatrix}$$



$$x_1=x_2=\pm\frac{1}{\sqrt{2}}$$

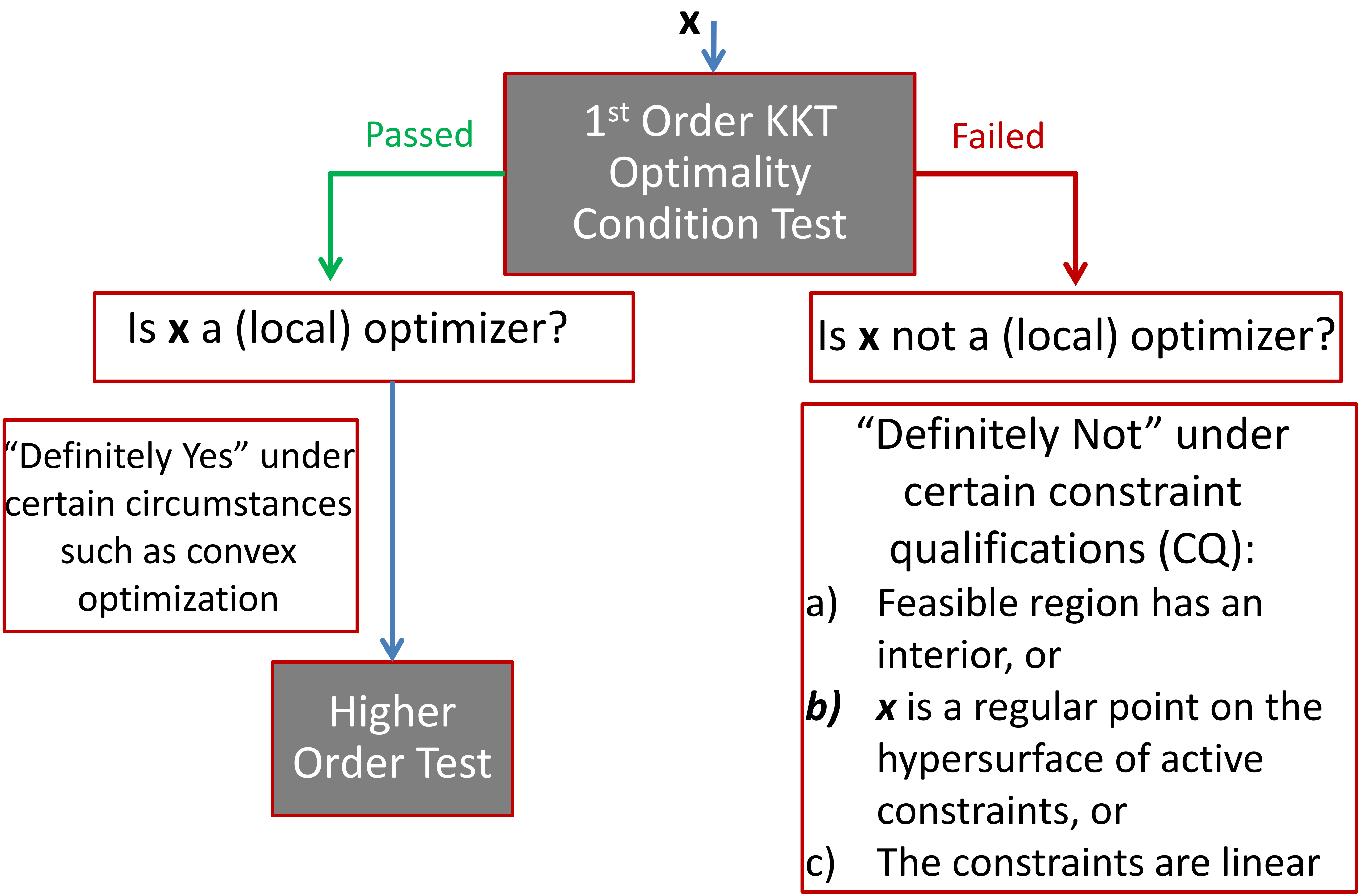
$$y=1\mp\sqrt{2}$$

$$x_1=x_2=\frac{1}{\sqrt{2}}$$

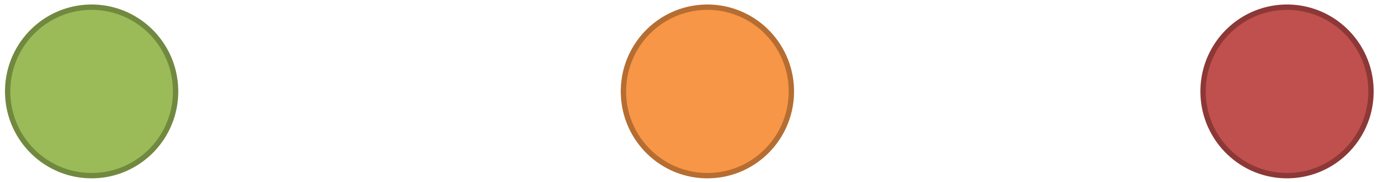
$$y=1-\sqrt{2}$$

Only the one with $y=1-2^{1/2}$ meets the SONC! (and it is also Sufficient)

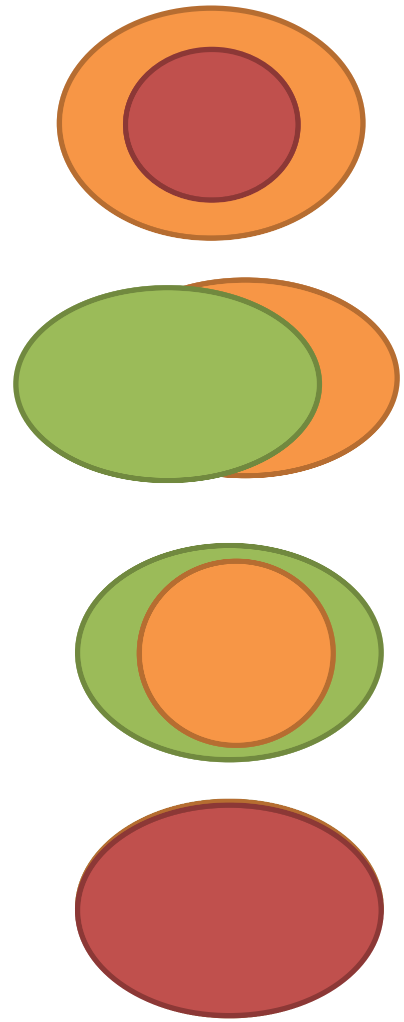
Summary of 1st Order KKT Optimality Test



Minimum and KKT Solutions



1st order KKT **Local Opt** **Global Opt**



Local vs Global Opt

KKT vs Local opt

KKT vs Local opt with CQ

KKT vs Global opt for CO with CQ