Chapter 4: Optimality Conditions

Introduction

- We now turn to the question of how to identify a local, i.e. also a global, minimum of a (convex) optimization problem.

- Reminder:

Local minimizer

A point x^* is a local minimizer if there is a neighborhood N of x^* such that $f(x*) \le f(x)$ for $x \in N$.

Global minimizer

A point x^* is the global minimizer if $f(x*) \le f(x)$ for all x from the domain of f

Convex Function

A function $f(x)$ is convex if for any two points x_1 and x_2, the function values satisfy inequality:

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2) \qquad for \qquad 0 \le \alpha \le 1$$

Active/inactive Inequality Constraint

Constraint for which does/does not determine the optimum

Outline

- Optimality Conditions for unconstrained problems
  - First-order necessary condition
  - Sufficient conditions

- Optimality Conditions for constrained problems
  - Lagrange multiplier theory
  - Necessary conditions (KKT)
  - Sufficient conditions

Necessary Conditions for Optimality (Unconstrained Problems)

- If the cost function is differentiable, we can use gradients and Taylor series expansions to copare the cost of a vector with the cost of ots close neighbours. In particular, we consider small variations Δx from a given vector x^*, which is approximately, up to first order, yoeld a cost variation

$$f(x^*+\Delta x) - f(x^*) \approx \nabla f(x^*)^T \Delta x$$

and up to second order, yield a cost variation

$$f(x^*+\Delta x) - f(x^*) \approx \nabla f(x^*)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^*) \Delta x$$

Approximation Using Taylor Expansion (1)

- Function of a single variable:

Consider a function ***f(x)*** of a single variable **x**

$$f(x) \approx f(x_0) + \frac{df}{dx}(x - x_0) + \frac{1}{2}\frac{d^2 f}{dx^2}(x - x_0)^2 + \cdots$$

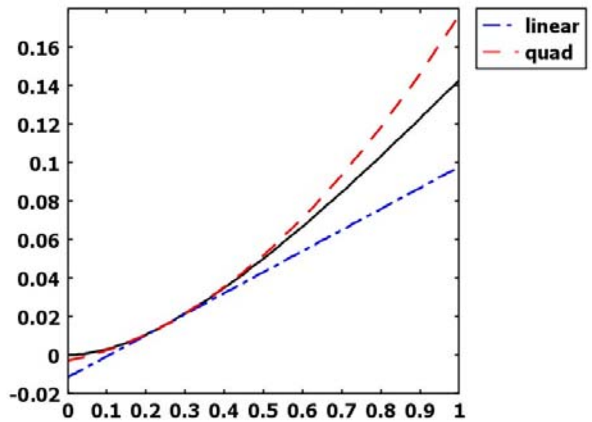
Linear approximation
Quadratic approximation

- Example: For function $f(x) = x^2/(4x+3)$

  and the given point $x_0 = 1/4$

  - Linear approximation
    $f(x) \approx 1/64 + 7/64(x - 1/4)$

  - Quadratic approximation
    $f(x) \approx 1/64 + 7/64(x - 1/4) + 9/64(x - 1/4)^2$



Approximation Using Taylor Expansion (2)

- Function of *N* variables

Consider a function $f(x)$ of n variables $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$ $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$

$$\boldsymbol{f}(x) \approx \boldsymbol{f}(\boldsymbol{x}_0) + \nabla\boldsymbol{f}^T\Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T\nabla^2\boldsymbol{f}\Delta\boldsymbol{x} + \cdots$$

------ Linear approximation

------ Quadratic approximation

- Example: For function $f = x^3 + 12xy^2 + 2y^2 + 5z^2 + xz^4$

  and the given point $x_0 = [1,2,3]$

  - Linear approximation

    $f \approx 475 + 132x + 56y + 138z$

  - Quadratic approximation

    $f \approx 535 - 294x + 3x^2 - 48y + 48xy + 14y^2 - 324z + 108xz + 59z^2$

$$\nabla\boldsymbol{f} = \begin{bmatrix} 3x^2 + 12y^2 + z^4 \\ 4(y + 6xy) \\ 2z(5 + 2xz^2) \end{bmatrix} \qquad \nabla^2\boldsymbol{f} = \begin{bmatrix} 6x & 24y & 4z^3 \\ 24y & 4 + 24x & 0 \\ 4z^3 & 0 & 2(5 + 6xz^2) \end{bmatrix}$$

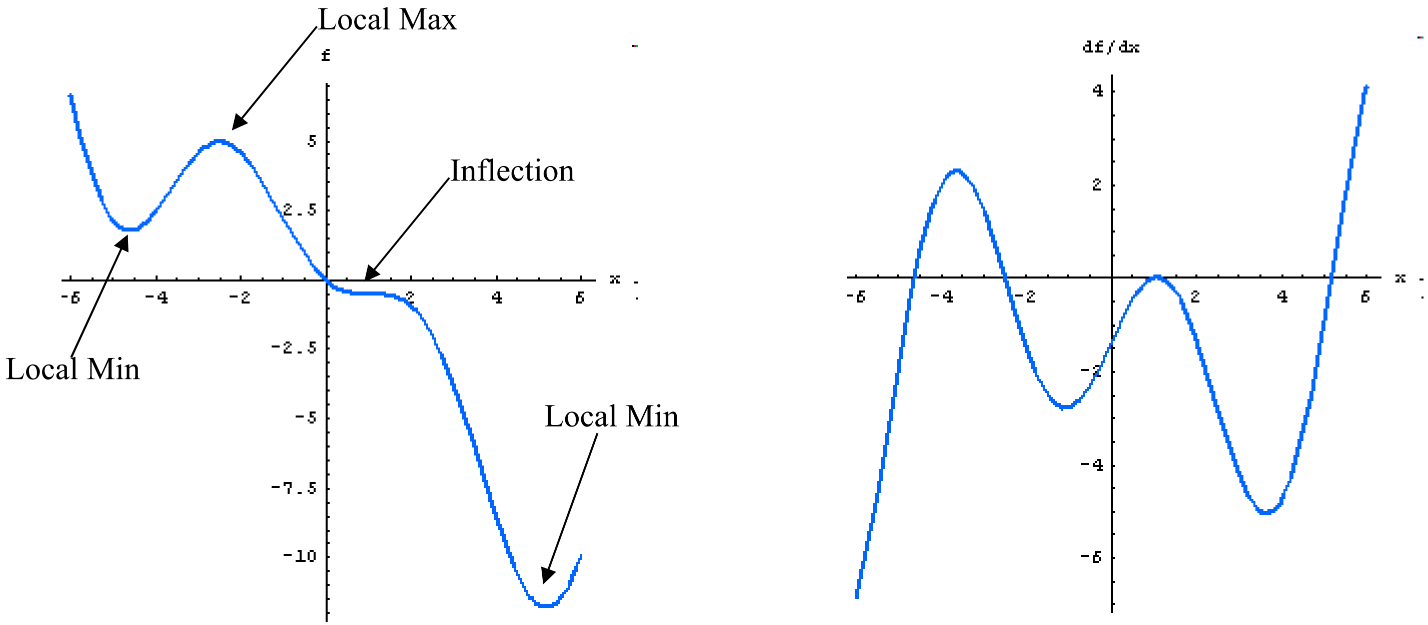
Optimality Conditions for Unconstrained Problems

Necessary Condition (First Order Condition)

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x} - \boldsymbol{x}^*) \qquad \nabla f \rightarrow \quad Gradient$$

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \approx \nabla f \cdot d \qquad d \rightarrow \quad StepSize$$

$$\nabla f(\boldsymbol{x}^*) = 0 \qquad \rightarrow \qquad \boldsymbol{x}^* \text{ is a (local) minimizer}$$



Necessary Conditions for Optimality (Unconstrained Problems)

We expect that if $x*$ is an unconstrained local minimun, the first order cost variation due to a small variation Δx is non negative:

$$\nabla f(x^*)^T \Delta x = \sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i} \Delta x_i \geq 0.$$

In particular, by taking Δx to be positive and negative multiples of the unit coordinate vectors (all coordinates equal to zero except or one which is equal to unity) we obtain

$$\partial f(x^*)/\partial x_i) \geq 0 \text{ and } \partial f(x^*)/\partial x_i) \leq 0$$

respectively, for all coordinates $i = 1, \ldots, n$. Equivalently, we have the necessary condition

$$\boxed{\nabla f(x^*) = 0}$$

stationary point

Illustration of the Necessary Optimality Conditions

- Figure shows how 1st order optimallity conditions can fail to guarantee local optimality of x^* if f is not convex.

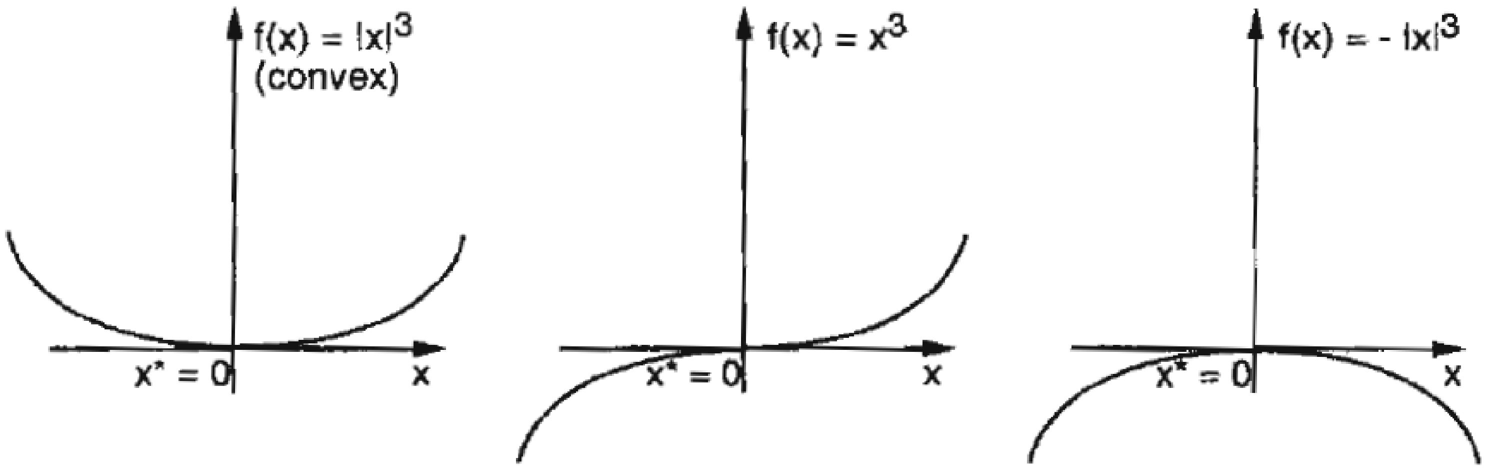


Illustration of the first order necessary optimality condition of zero slope $[\nabla f(x^*) = 0]$ for funtions of one variable.

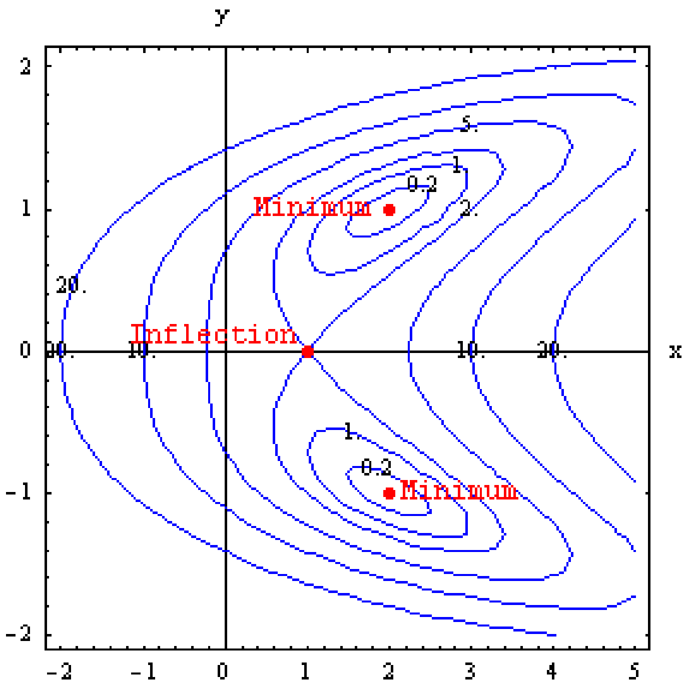
- The first order conditionis satisfied not only by local minima, but also by local maxima and „inflection" points, such as the one in the middle figure above.
- If the function is convex, the condition $\nabla f(x^*) = 0$ necessary and sufficient for global optimality of x^*.

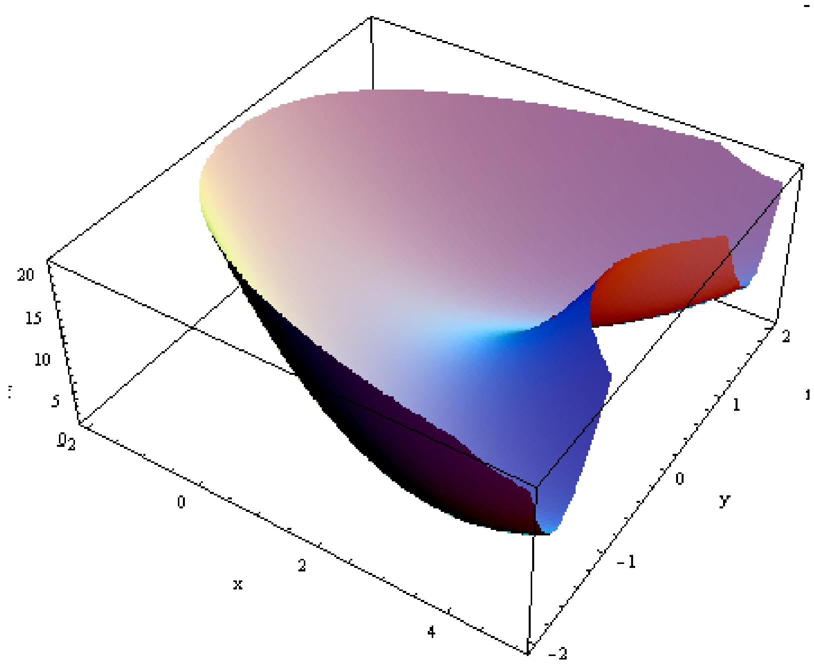
Source: Fig. 1.1.3 D. P. Bertsekas, Nonlinear Programming

Remarks

- Necessary Condition only tell us that objective function value in the small neighbourhood of x^* is not increasing.
- Necessary Condition is true even for a maximum point and inflection point (where locally the gradient is also zero).
- All the points which satisfy the necessary condition are called Stationary Points.

$$f = (x-2)^2 + (x-2y^2)^2$$





Sufficient Conditions for Optimality (Unconstrained Problems)

- Suppose we have a vector x^* that satisfies the first order necessary optimality condition

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

and also satisfies the following strengthened form of the second order necessary optimality condition

$$\nabla^2 f(x^*): \text{positive definite}$$

That is the Hessian is positive definite. Then, for all $\Delta x \neq 0$ we have (according to the 2nd order Taylor approximation:

$$\Delta x' \nabla^2 f(x^*) \Delta x > 0$$

implying that at x^* the second order variation of f due to a small nonzero variation Δx is positive.

Sufficient Conditions for Optimality (Unconstrained Problems)

$$f(x) \approx f(x^*) + \nabla f^T \cdot d + \frac{1}{2} d^T \nabla^2 f \cdot d$$

←Second-order Taylor expansion

$$f(x) - f(x^*) \approx \nabla f^T \cdot d + \frac{1}{2} d^T \nabla^2 f \cdot d$$

From the necessary condition of optimization $\nabla f(x^*) = 0 \quad \Rightarrow$

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq 0 \quad \text{when} \quad \frac{1}{2} d^T \nabla^2 f \cdot d \geq 0$$

$d^T \nabla^2 f \cdot d > 0$ when $\nabla^2 f(x^*)$ is positive definite x^* is ***minimum point***

$d^T \nabla^2 f \cdot d < 0$ when $\nabla^2 f(x^*)$ is negative definite x^* is ***maximum point***

$d^T \nabla^2 f \cdot d \geq 0$ when $\nabla^2 f(x^*)$ is positive semidefinite,
then the second derivative test is ***inconclusive***

First Derivative Test

- In calculus, the first derivative test uses the first derivative of a function to determine whether a given stationary point of a function is a local maximum, a local minimum, or neither.
- The first derivative test depends on the "increasing-decreasing test", which is itself ultimately a consequence of the Lagrange theorem (see Excursus Analysis).

- Suppose f is a real-valued function of a real variable defined on some interval containing the critical point x*. Further suppose that f is continuous at x and differentiable on some open interval containing x*, except possibly at x* itself.

  - If there exists a positive number r such that for every y in $(x - r, x)$ we have $f'(y) \geq 0$, and for every y in $(x, x + r)$ we have $f'(y) \leq 0$, then f has a local maximum at x.
  - If there exists a positive number r such that for every y in $(x - r, x)$ we have $f'(y) \leq 0$, and for every y in $(x, x + r)$ we have $f'(y) \geq 0$, then f has a local minimum at x.
  - If there exists a positive number r such that for every y in $(x - r, x) \cup (x, x + r)$ we have $f'(y) > 0$, or if there exists a positive number r such that for every y in $(x - r, x) \cup (x, x + r)$ we have $f'(y) < 0$, then f has neither a local maximum nor a local minimum at x.
  - If none of the above conditions hold, then the test fails. (there are functions that satisfy none of the first three conditions.)

Second Derivative Test

- In calculus, the second derivative test is a criterion often useful for determining whether a given stationary point of a function is a local maximum or a local minimum using the value of the second derivative at the point.

- The test states: If the function f is twice differentiable at a stationary point x^* ($f'(x^*) = 0$), then:

  - if $f''(x^*) > 0$ then has a local maximum at x^*.
  - if $f''(x^*) < 0$ then has a local minimum at x^*.
  - if $f''(x^*) = 0$, the second derivative test says nothing about the point x^*, a possible inflection point.

- In the last case, although the function may have a local maximum or minimum at x*, because the function is sufficiently "flat" (i.e. $f''(x^*) = 0$) the extremum is rendered undetected by the second derivative. In this case one has to examine the third derivative. The point at which $f''(x^*) = 0$ is an inflection point if convexity changes on either side of it.

Proof of the Second Derivative Test

Suppose we have $f''(x) > 0$ (the proof for $f''(x) < 0$ is analogous). Then:

$$0 < f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h \to 0} \frac{f'(x+h) - 0}{h} = \lim_{h \to 0} \frac{f'(x+h)}{h}.$$

thus, for h sufficiently small we get

$$\frac{f'(x+h)}{h} > 0$$

which means that $f'(x+h) < 0$ if $h < 0$ so that f is decreasing to the left of x,

and that $f'(x+h) > 0$ if $h > 0$ so that f is increasing to the right of x.

now, by first derivative test, we know that f has a local minimum at x.

Convexity Test

- The second derivative test may also be used to determine the convexity of a function as well as a function's points of inflection (see chapter 3).

- First, all points at which $f''(x) = 0$ are found. In each of the intervals created, f"(x) is then evaluated at a single point. For the intervals where the evaluated value of f"(x) < 0 the function f(x) is concave, and for all intervals between critical points where the evaluated value of f"(x) > 0 the function f(x) is convex. The points that separate intervals of opposing convexity are points of inflection (saddle points).

- For example, (0,0) is an inflection point on $f(x) = x^3$ because $f''(0) = 0$, $f''(-1) < 0$, and $f''(1) > 0$.

- Caution: A function can also change the convexity in a point in which it is not differentiable!

Finding the Optimum Using Optimality Conditions

- The most straight forward method to use optimality conditions to solve an optimization problem, is as follows: First, find all points satisfying the first order necessary condition $\nabla f(x) = 0$; then (if f is not known to be convex), check the second order necessary condition(∇^2f is positive semidefinite) for each of these points, filtering out those that do not satisfy it; finally for the remaining candidates, check if ∇^2f is positive definite, in which case we are sure that they are strict local minima.

Example: Function with One Minimum, Two Local Maximum and Two Inflection Points

Gradient vector → $\begin{pmatrix} 50x - 48x^3 - 6y - 48xy^2 \\ -6x + 50y - 48x^2y - 48y^3 \end{pmatrix}$

Necessary conditions → $\begin{pmatrix} 50x - 48x^3 - 6y - 48xy^2 == 0 \\ -6x + 50y - 48x^2y - 48y^3 == 0 \end{pmatrix}$

stationary points → $\begin{pmatrix} x \to -0.763763 & y \to 0.763763 \\ x \to -0.677003 & y \to -0.677003 \\ x \to 0. & y \to 0. \\ x \to 0.677003 & y \to 0.677003 \\ x \to 0.763763 & y \to -0.763763 \end{pmatrix}$

Hessian matrix → $\begin{pmatrix} 50 - 144x^2 - 48y^2 & -6 - 96xy \\ -6 - 96xy & 50 - 48x^2 - 144y^2 \end{pmatrix}$

Point → {x → -0.763763, y → 0.763763} MaximumPoint → 16.33

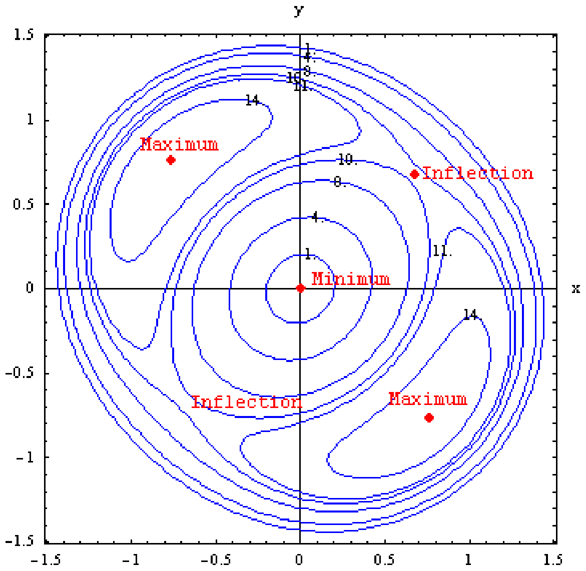
Hessian → $\begin{pmatrix} -62. & 50. \\ 50. & -62. \end{pmatrix}$

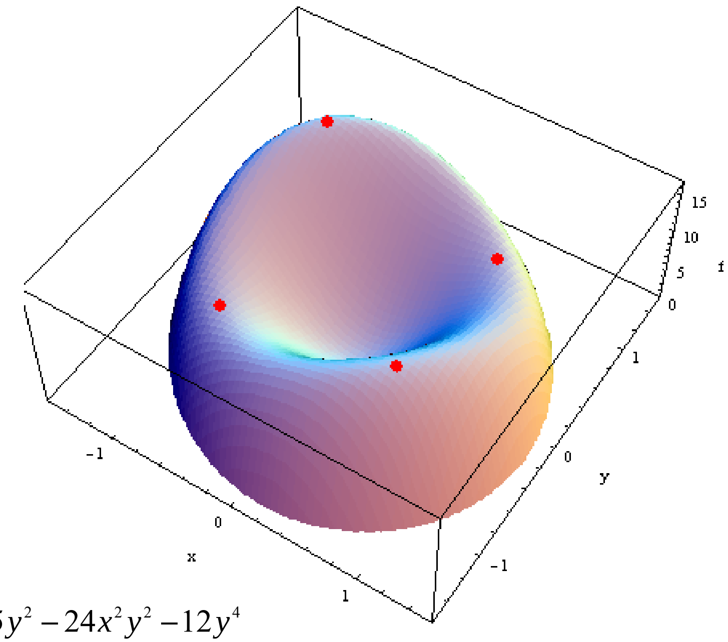
Point → {x → 0., y → 0.} MinimumPoint → 0

Hessian → $\begin{pmatrix} 50. & -6. \\ -6. & 50. \end{pmatrix}$

Point → {x → 0.677003, y → 0.677003} InflectionPoint → 10.08

Hessian → $\begin{pmatrix} -38. & -50. \\ -50. & -38. \end{pmatrix}$





$$f = 25x^2 - 12x^4 - 6xy + 25y^2 - 24x^2y^2 - 12y^4$$

Summary for Unconstrained Problems

If $f(\boldsymbol{x}^*), \nabla f(\boldsymbol{x}^*), \nabla^2 f(\boldsymbol{x}^*)$ is continuously differentiable in an open neighbourhood of x^* point

- First-order Necessary Condition $\nabla f(\mathrm{x}^*) = 0$
- Second-order Sufficient Condition $\nabla f(\mathrm{x}^*) = 0$,$\nabla^2 f(\mathrm{x}^*)$ is positive definite
- Since there is no constraint, the optimization problem is a convex problem if the objective function is convex.

Optimality Conditions for Constrained Problems (Equality Constraints)

- Consider the two-dimensional optimization problem with a single equality constraint:

$$\max f(x,y)$$
$$\text{s.t.}\quad g(x,y)=c$$

- First approach is the ***elimination method***, that is to express one variable as a function of the other one and than substitute in f(x,y), which becomes a function of a single variable.
  - **Example:** A can producer wants to maximize a volume of his cans. The amount of material (which corresponds to the can surface) should be A0. What are the optimal values of can diameter and height?
  - **Solution:** Objective function is a can volume:
    whereas its surface defines an equality constraint $V(r,l)=\pi r^2 l,$
    If we express *l* as a function of r and A0 and substitute i $A(r,l)=2\pi r^2+2\pi r l=A_0.$
    get:

    $$V_1(r)=V\left(r,\frac{A_0}{2\pi r}-r\right)=\frac{A_0}{2}r-\pi r^2.$$

    Setting

    $$r^{\text{opt}}=A_0/6\pi.$$

- Unfortunately, the substitution method often leads to analytically very complicated expressions and hence, in majority of problems is not practical. An alternative is a ***Lagrange multipliers method***.

Lagrange Multipliers (I)

- In mathematical optimization, the method of Lagrange multipliers (named after Joseph Louis Lagrange) provides a strategy for finding the maxima and minima of a function subject to ***equality constraints***.
- Consider the two-dimensional problem introduced above:

$$\max f(x,y)$$
$$\text{s.t.} \quad g(x,y) = c$$

We can visualize contours of f given by $f(x,y) = d$, for various values of d, and the contour of g given by $g(x,y) = c$.

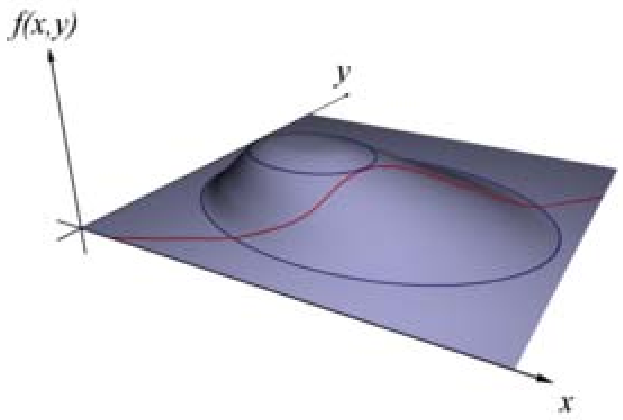


Figure 1: Find x and y to maximize $f(x,y)$ subject to a constraint (shown in red) $g(x,y) = c$.

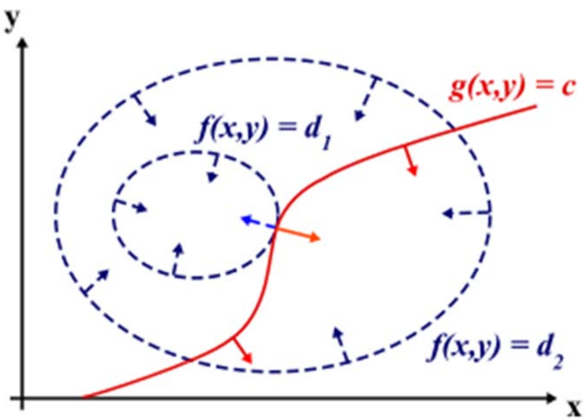
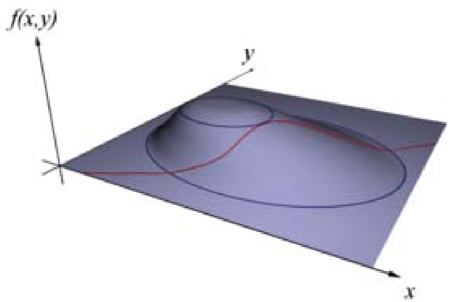
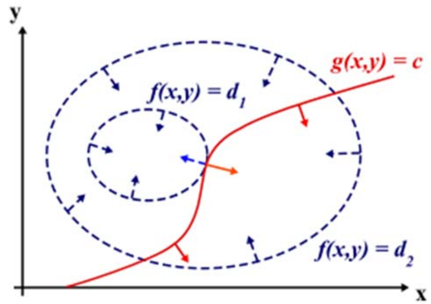


Figure 2: Contour map of Figure 1. The red line shows the constraint $g(x,y) = c$. The blue lines are iso-contours of $f(x,y)$. The point where the red line tangentially touches a blue contour is our solution.

Source: Wikipedia

Lagrange Multipliers (II)





Source: Wikipedia

- Suppose we walk along the contour line with $g = c$. In general the contour lines of f and g may be distinct, so following the contour line for $g = c$ one could intersect with or cross the contour lines of f. This is equivalent to saying that while moving along the contour line for $g = c$ the value of f can vary. Only when the contour line for $g = c$ meets contour lines of f tangentially, do we not increase or decrease the value of f , that is, when the contour lines touch but do not cross.
- The contour lines of f and g touch when the tangent vectors of the contour lines are parallel. Since the gradient of a function is perpendicular to the contour lines (iso-lines), this is the same as saying that **the gradients of f and g are parallel**. Thus we want points (x,y) where:

$$\nabla f(x,y) = -\lambda \nabla g(x,y)$$

with λ being an arbitrary scalar and gradients defined as: $\nabla f(x,y) = \left[\frac{\partial f}{\partial x} \;\; \frac{\partial f}{\partial y}\right]^T, \quad \nabla g(x,y) = \left[\frac{\partial g}{\partial x} \;\; \frac{\partial g}{\partial y}\right]^T.$

Lagrange Multipliers (III)

- To incorporate these conditions into one equation, we introduce an auxiliary function

$$\Lambda(x, y, \lambda) = f(x, y) - \lambda \cdot (g(x, y) - c)),$$

called ***Lagrangian function*** and solve:

$$\frac{\partial \Lambda}{\partial x} = 0;$$

$$\frac{\partial \Lambda}{\partial y} = 0;$$

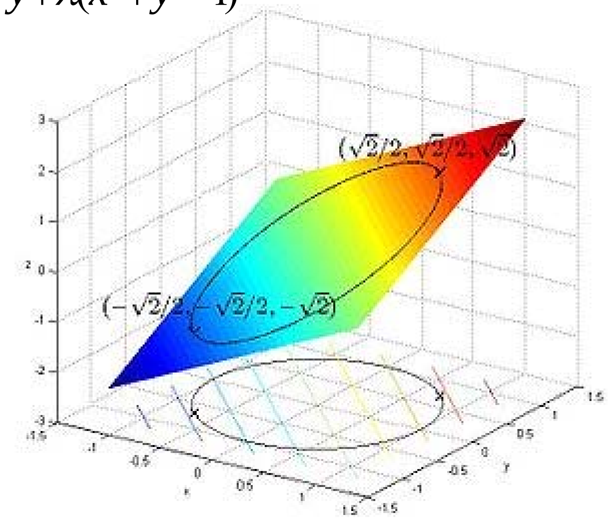
$$\frac{\partial \Lambda}{\partial \lambda} = 0;$$

- This is the method of Lagrange multipliers. Note that the solution of the third equation implies the given constraint $g(x,y) = c$.
- Lagrange Multipliers method delivers only the ***necessary condition*** for constrained optima of the objective function, i. e. not all stationary points of Lagrangian function must be contrained optima of the objective function!

Example

- Suppose one wishes to maximize $f(x,y) = x + y$ subject to the constraint $x^2 + y^2 = 1$. The feasible set is the unit circle, and the level sets of f are diagonal lines (with slope -1), so one can see graphically that the maximum occurs at $\left(\sqrt{2}/2, \sqrt{2}/2\right)$, and the minimum occurs at $\left(-\sqrt{2}/2, -\sqrt{2}/2\right)$.

Formally, set $g(x,y) - c = x^2 + y^2 - 1$,and $\Lambda(x,y,\lambda) = f(x,y) + \lambda(g(x,y) - c) = x + y + \lambda(x^2 + y^2 - 1)$



Set the derivations $d\Lambda = 0$, which yields the system of equations:

$$\frac{\partial \Lambda}{\partial x} = 1 + 2\Lambda x = 0 \rightarrow (i)$$

$$\frac{\partial \Lambda}{\partial y} = 1 + 2\Lambda y = 0 \rightarrow (ii)$$

$$\frac{\partial \Lambda}{\partial \lambda} = x^2 + y^2 - 1 = 0 \rightarrow (iii)$$

As Always, the $\partial\lambda$ equation ((iii) here) is the original constraint.

Combinig the first two equations yields x=y (explicitly, $\lambda \neq 0$ otherwise(i) yields 1=0, so one has

Substituting into (iii) yields $2x^2 = 1$, so $x = y = \pm\sqrt{2}/2 \,\&\, \lambda = \mp\sqrt{2}/2$ showing the stationary points are $(\sqrt{2}/2, \sqrt{2}/2) \,\&\, (-\sqrt{2}/2, -\sqrt{2}/2)$

Evaluating the objective function f on these yields $f(\sqrt{2}/2, \sqrt{2}/2) = \sqrt{2} \;\&\; f(-\sqrt{2}/2, -\sqrt{2}/2) = -\sqrt{2}$

thus the maximum is $\sqrt{2}$ which is attained at $(\sqrt{2}/2, \sqrt{2}/2)$ and the minimun is $-\sqrt{2}$ which is attained at $(-\sqrt{2}/2, -\sqrt{2}/2)$

Source: Wikipedia

Multiple Equality Constraints

- The method of Lagrange multipliers can also accommodate multiple constraints:

$$\begin{aligned} &\min\ f(x_1,\ldots,x_n) \\ &s.t.\ \ g_1(x_1,\ldots,x_n)=0 \\ &\qquad g_2(x_1,\ldots,x_n)=0 \\ &\qquad\qquad \vdots \\ &\qquad g_m(x_1,\ldots,x_n)=0 \end{aligned}$$

- We introduce a vector of Lagrange multipliers $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\ldots,\lambda_m]^T$ and define the Lagrangian function as:

$$\Lambda(x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_m) = f(x_1,\ldots,x_n) - \sum_{k=1}^{m} \lambda_k g_k(x_1,\ldots,x_n),$$

and solve $\nabla_{x_1,\ldots,x_n,\lambda_1,\ldots\lambda_m} \Lambda = 0$, that is

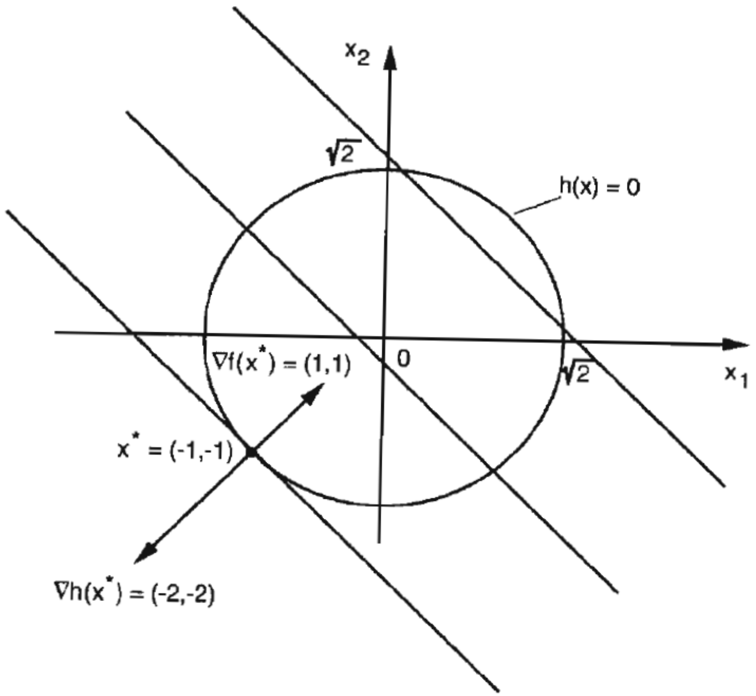
$$\nabla f(x_1,\ldots,x_n) - \sum_{k=1}^{m} \lambda_k \nabla g_k(x_1,\ldots,x_n) = 0$$

$$g_{k=1,\ldots m}(x_1,\ldots,x_n) = 0$$

Interpretation of Lagrange Multiplier Condition

- A condition $\nabla f(x^*) + \sum_{i=1}^{m} \lambda \nabla g_i(x^*) = 0$ indicates that the cost gradient $\nabla f(x^*)$ at a given local optimum x^* belongs to the subspace spanned by the constraint gradients at x^*.

- Example: minimize $x_1 + x_2$
  subject to $x_1{}^2 + x_2{}^2 = 2$



At the local minimum $x^* = (-1, -1)$, the cost gradient $\nabla f(x^*)$ is normal to the constraint surface and is therefore, colinear with the constraint gradient $\nabla h(x^*) = (-2, -2)$. The Lagrange multiplier is $\lambda = 1/2$.

Source: Fig. 3.1.1 D. P. Bertsekas, Nonlinear Programming

Formal Statement of the Lagrange Multiplier Theorem (Necessary Condition for Equality Constraints)

- LANGRANGE MULTIPLIER THEOREM – NECESSARY CONDITIONS: Let x^* be a local minimum of f subject to $h(x) = 0$ and assume that the constraint gradients $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent. Then there exists a unique vector $\lambda^* = (\lambda_1{}^*, \ldots, \lambda_m{}^*)$ called a Lagrange multiolier vector such that

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

- For easy reference, a feasible vector x for which the constraint gradients $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent will be called **regular**.

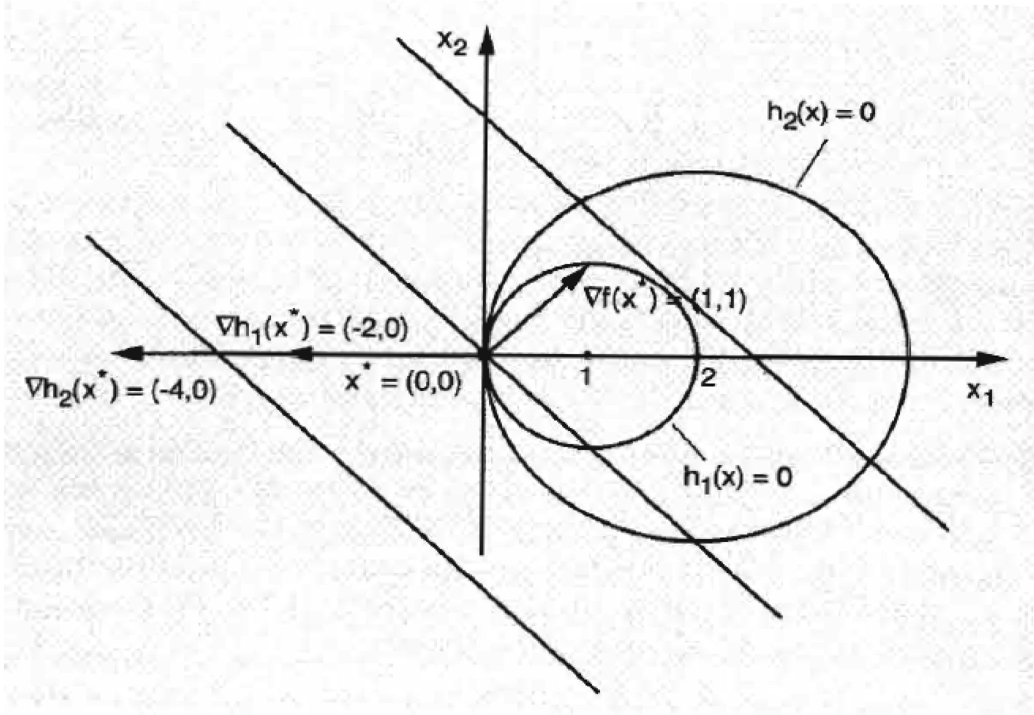
- We will see soon that there may not exist Lagrange multipliers for a local minimum that is not regular.

A Problem with no Lagrange Multiplier

- Minimize $f(x) = x_1 + x_2$

  subject to $h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0$

  $h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0$



At the local minimum $x^* = (0,0)$ (the only feasible solution), the cost gradient $\nabla f(x^*) = (1,1)$ cannot be expressed as a linear combination of the constraint gradients $\nabla h_1(x^*) = (-2,0)$ and $\nabla h_2(x^*) = (-4,0)$. Thus the Lagrange multiplier condition

$$\nabla f(x^*) + \lambda_1^* \nabla h_1(x^*) + \lambda_2^* \nabla h_2(x^*) = 0$$

cannot hold for any λ_1^* and λ_2^*.

Source: Fig. 3.1.2 D. P. Bertsekas, Nonlinear Programming

Optimality Conditions for Constrained Problems (Inequality Constraints)

- We now turn to the question of how to identify a local, i.e. also a global, minimum of a (convex) optimization problem with inequality constraints.

$$
\text{(ICP)} \quad \begin{cases} \min f(\mathbf{x}) \\ \text{s.t.} \begin{cases} g_j(\mathbf{x}) \leq 0, & j = 1, \cdots, r \\ h_i(\mathbf{x}) = 0, & i = 1, \cdots, m \end{cases} \end{cases}
$$

- To that end we first define the Lagrangian function:

$$
L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x).
$$

Active/Inactive Inequality Constraints

- For any feasible point x, the set of active inequality constraints is denoted by

$$A(x) = \{ j \mid g_j(x) = 0 \}$$

- If $j \notin A(x)$, we say that the jth constraint is inactive at x. We note that if x^* is local minimum of the inequality constrained problem (ICP) then x^* is also local minimum for a problem identical to (ICP) except that the inacive constraints at x^* have been discarded. Thus in effect, inactive constraints at x^* don‘t matter; they can be ignored in the statement of optimality conditions.

- On the other hand, at a local mimimum, active inequality constraints can be treated to a large extent as equalities. In particular, if x^* is a local minimum of the inequality constrained problem (ICP) then x* is also a local minimum for the equality constrained problem

$$\begin{cases} \min f(\mathbf{x}) \\ \text{s.t. } h_1(\mathbf{x}) = 0, \ldots, h_m(\mathbf{x}) = 0, \quad g_j(\mathbf{x}) = 0, \forall j \in A(x^*) \end{cases}$$

Active/Inactive Inequality Constraints

- Thus if x* is regular for the letter problem, there exist Lagrange multipliers $\lambda_1{}^*, \dots, \lambda_m{}^*$ and $\mu_j{}^*, j \in A(x^*)$ such that

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

- Assigning zero Lagrange multipliers to the inactive constraints, we obtain

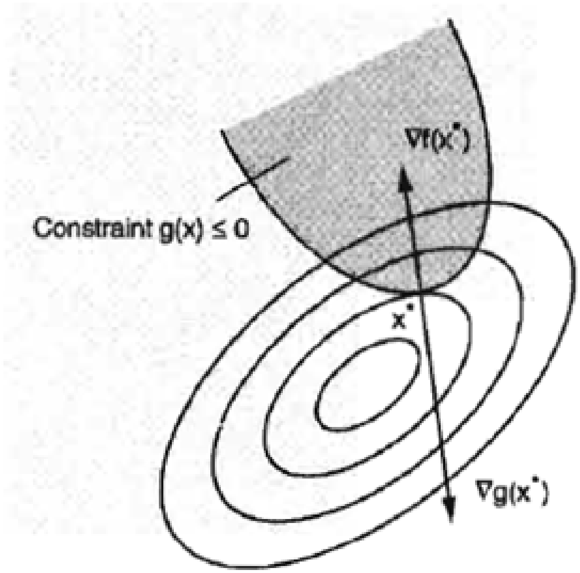
$$\boxed{\begin{aligned} &\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0 \\ &\mu_j^* = 0, \forall j \notin A(x^*) \end{aligned}}$$

which may be viewed as an analog of the first order optimality condition for the equality constrained problem.

Nonnegative Lagrange Multiplier

- There is one important fact about the Lagrange multipliers μ_j^*, they are non negative.

- Example: The figure shows the illustration of the non negativity of the Lagrange multiplier for a problem with a single inequality constraint. If the constraint is inactive, then $\mu^* = 0$. Otherwise, $\nabla f(x^*)$ is normal to the constraint surface and the points to the inside of the constraint set, while $\nabla g(x^*)$ is normal to the constraint surface and the points to the outside of the constraint set. Thus $\nabla f(x^*)$ and $\nabla g(x^*)$ are collinear and have opposite signs, implying that the Lagrange multiplier is non negative.



Source: Fig. 3.3.1 D. P. Bertsekas, Nonlinear Programming

Formal Statement of the Karush-Kuhn-Tucker Theorem
(Necessary Conditions for Inequality Constraints)

- KARUSH-KUHN-TUCKER THEOREM NECESSARY CONDITIONS: Let x^* be a local minimum of the problem

$$\begin{cases} \min\ f(x) \\ \text{s.t.} \begin{cases} h_1(x)=0,\ldots\ldots,h_m(x)=0 \\ g_1(x)\le 0,\ldots\ldots,g_r(x)\le 0 \end{cases} \end{cases}$$

Where f, h_i, g_j are continuously differentiable functions from R^n to R, and assume that x^* is regular. Then there exist unique Lagrange multiplier vectors $\lambda^* = (\lambda_1{}^*, \ldots, \lambda_m{}^*)$, $\mu^* = (\mu_1{}^*, \ldots, \mu_r{}^*)$ such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$
$$\mu_j^* \ge 0,\ j = 1\ldots.r,$$
$$\mu_j^* = 0 \forall j \notin A(x^*)$$

where A(x^*) is the set of active constraints at x^*.

KKT are Only the Necessary Conditions

- For sufficiently regular nonconvex problems, the KKT conditions are necessary, but not sufficient, optimality conditions for (P).
- That is, local optima are always found among the KKT points, but there may be KKT points that are not local optima.
- The fact that the KKT conditions cannot be sufficient for optimality is evident by studying the special case of an unconstrained optimization problem, where the KKT points are equivalent to stationary points.
- Numerical algorithms typically try to find KKT points, and thus one may end up at a point that is not a local minimum, but even a local maximum! However, for convex problems a KKT point is always an optimal point.

The Regularity Condition

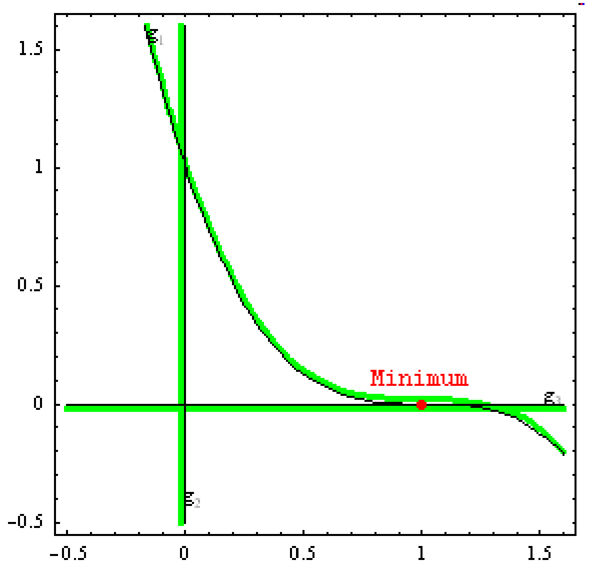
- Active inequality and all equality constraints must be linearly independent.

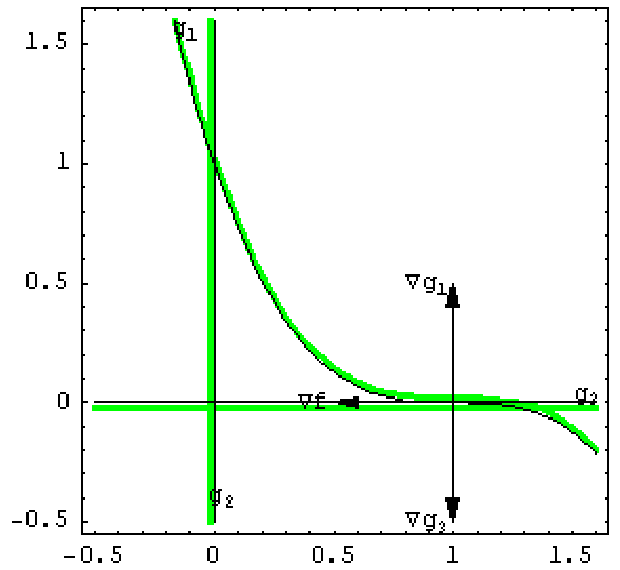
- The optimality conditions make sense only at points that are regular.

- The KKT condition $\nabla f(x^*) = -\sum_{i \in Active} \alpha_i \nabla g_i(x^*) + \sum_{i=1}^{p} \beta_i \nabla h_i(x^*)$ may not work for an irregular point, that can be minimum point.

$$f(x, y) = -x$$

$$g = \begin{Bmatrix} y - (1-x)^3 \le 0 \\ -x \le 0 \\ -y \le 0 \end{Bmatrix}$$



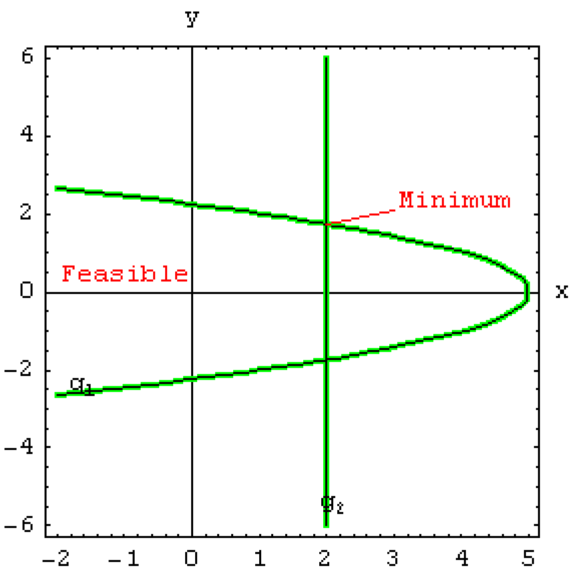


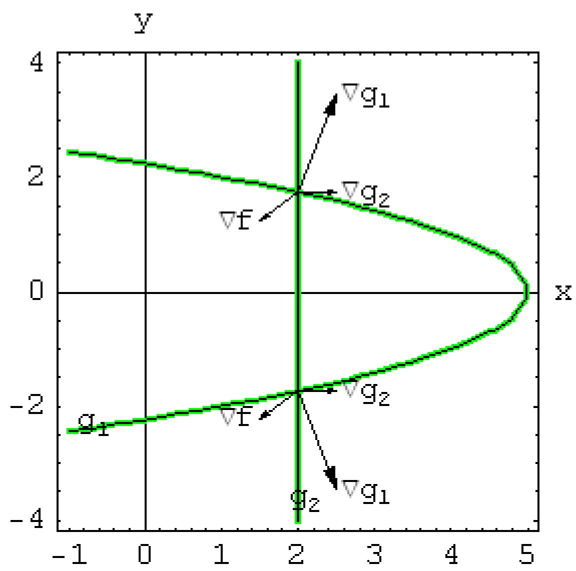
Example

For problems with two variables, it is possible to draw gradient vectors and geometrically interpret the KKT condition

$$f = -x - y$$
$$g = \{x + y^2 - 5 \le 0; x - 2 \le 0\}$$

$$\nabla f \rightarrow \begin{pmatrix} -1 \\ -1 \end{pmatrix} \qquad \nabla g_1 \rightarrow \begin{pmatrix} 1 \\ 2y \end{pmatrix} \quad \nabla g_2 \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$





Geometrical Interpretation of KKT Condition

- The KKT conditions state that $-\nabla f(x^*)$ should belong to the cone Spanned by the gradients of the active Constraints at a point x^*.

- Example: At point $\overline{x1}$ in the figure,

$$-\nabla g_0(\bar{x}_1) = \lambda_1 \nabla g_1(\bar{x}_1) + \lambda_2 \nabla g_2(\bar{x}_1), \lambda_1 \geq 0, \lambda_2 \geq 0$$

so, $\bar{x}_1$ is a KKT point and consequently the optimal solution. Regarding point $\bar{x}_2$, we see that $-\nabla g_0(\bar{x}_2)$ does not belong to the cone spanned by the active constraints at $\bar{x}_2$. That is there does not exist any $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$ such that

$$-\nabla g_0(\bar{x}_2) = \lambda_2 \nabla g_2(\bar{x}_2) + \lambda_3 \nabla g_3(\bar{x}_2)$$

and consequently $\bar{x}_2$ is not a KKT point.

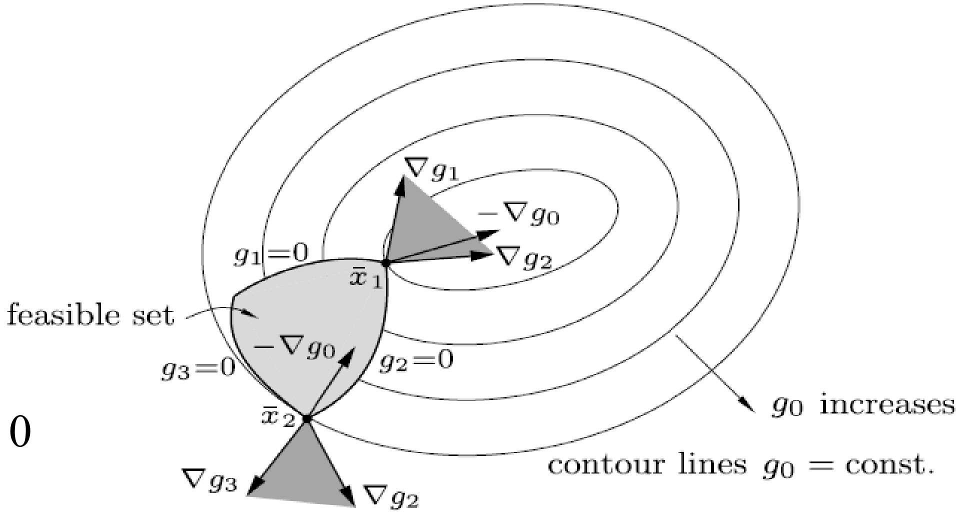


Illustration of the KKT conditions

Source: Fig. 3.5 W. Christensen, A. Klabring, An Introduction to Structural Optimization

Geometrical Interpretation of KKT Condition

Case 1 : No active constraint $\Rightarrow$ $\nabla f(\boldsymbol{x}^*) = 0$

Case 2 : One constraint **:** $\Rightarrow$ $\nabla f(\boldsymbol{x}^*), \nabla g(\boldsymbol{x}^*)$ lie along the same line but opposite in direction. The Lagrangian multiplier is simply a scale factor

Case 3 : Multiple constraint **:** $\Rightarrow$ $\nabla f(\boldsymbol{x}^*)$ and the $\alpha \nabla g_i^T(\boldsymbol{x}^*) + \beta \nabla h_j^T(\boldsymbol{x}^*)$ lie along the same line but opposite in direction. The parameter α, β are simply scale factors

Second-order Sufficiency Conditions

- Except for convex problem, the KKT condition are only the *Necessary Condition* for the minimum.

- Since the Lagrangian function is essentially unconstrained, the sufficient condition that a given KKT point, is a local minimum reads:

$$\boldsymbol{d}^T[\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=active} \alpha_i \nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \beta_j \nabla^2 h_j(\boldsymbol{x}^*)]\boldsymbol{d} > 0$$

Remarks

- Optimality conditions provide powerful means to verify solutions
- But there are limitations…
  - Sufficiency conditions are difficult to verify.
  - Practical problems do not have required nice properties.
  - May have problems if you do not know the explicit constraint equations (e.g., in FEM).

Homework

- Solve the problem of weight minimization of a two-bar truss subject to stress and displacement constraints (problem 2.2 from lecture 2) usig KTT conditions and drow gradients at the solution.