Optimizasyon

(En İyileme)

Dr. Cahit Karakuş

Optimizasyon

- Matematiksel modeller, bir mühendislik probleminin çözümüne ulaşmak için ilgilendiğimiz sistemin veya makinanın davranışını simüle etmek için kullandığımız araçlardı;
- Optimizasyon ise problemin birden fazla çözümü olduğunda en iyi çözümü üretme çabasıdır.

Mühendislik Uygulamaları Açısından Optimizasyonun Temel Unsurları

- Problemin, hedefini içeren bir ***amaç fonksiyonu*** olacaktır.
- Bir takım ***tasarım değişkenleri*** olacaktır. Bu değişkenler reel veya tamsayı olabilirler.
- Problemde çalıştığımız sınırlayıcı koşulları tanımlayan ***kısıtlar*** olacaktır.

Özdeğer ve Özvektörlerin Geometrik Yorumu

Bir matris herhangi bir vektör ile çarpıldığında yeni bir vektör elde edilir. Bu durumda bir matris bir sistemi ve bir vektör bir sistemdeki bir sinyalin zaman içerisindeki değişimini temsil ediyorsa. Bir matris ile bir vektör çarpıldığında matris vektörün yönünü, doğrultusunu ve şiddetini değiştirir. Vektörün davranışı değişir.

$$\begin{pmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \\ x3 \end{pmatrix} = \begin{pmatrix} b1 \\ b2 \\ b3 \end{pmatrix}$$

$$\begin{pmatrix} b1 \\ b2 \\ b3 \end{pmatrix} = \begin{pmatrix} a11 * x1 + a12 * x2 + a13 * x3 \\ a21 * x1 + a22 * x2 + a23 * x3 \\ a31 * x1 + a32 * x2 + a33 * x3 \end{pmatrix}$$

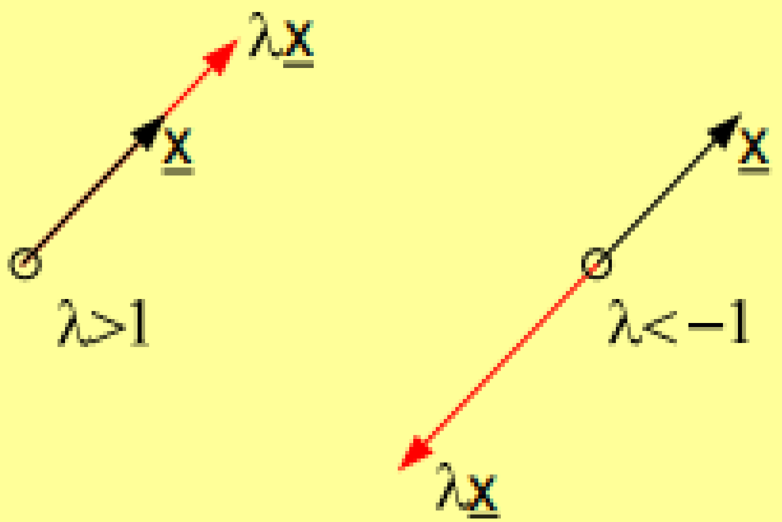
Bazı vektörler bir A matrisi ile çarpıldıkları zaman vektörün doğrultusu değişir, yönü değişmez. Bu özel x vektörleri, Ax vektörü ile aynı doğrultuda ya da ters doğrultuda kalmaktadır. İşte bu vektörlere "özvektörler" denir. Özdeğerler, bir matrisin orijinal yapısını görmek için kullanılan alternatif bir yoldur. A matrisi x vektörünü özdeğer kadar büyütmekte ya da küçültmektedir. Özdeğer pozitif ise her iki vektör aynı yönde, özdeğer negatif ise vektörler ters yöndedir. (Quantum …)

Özdeğer, Özvektör

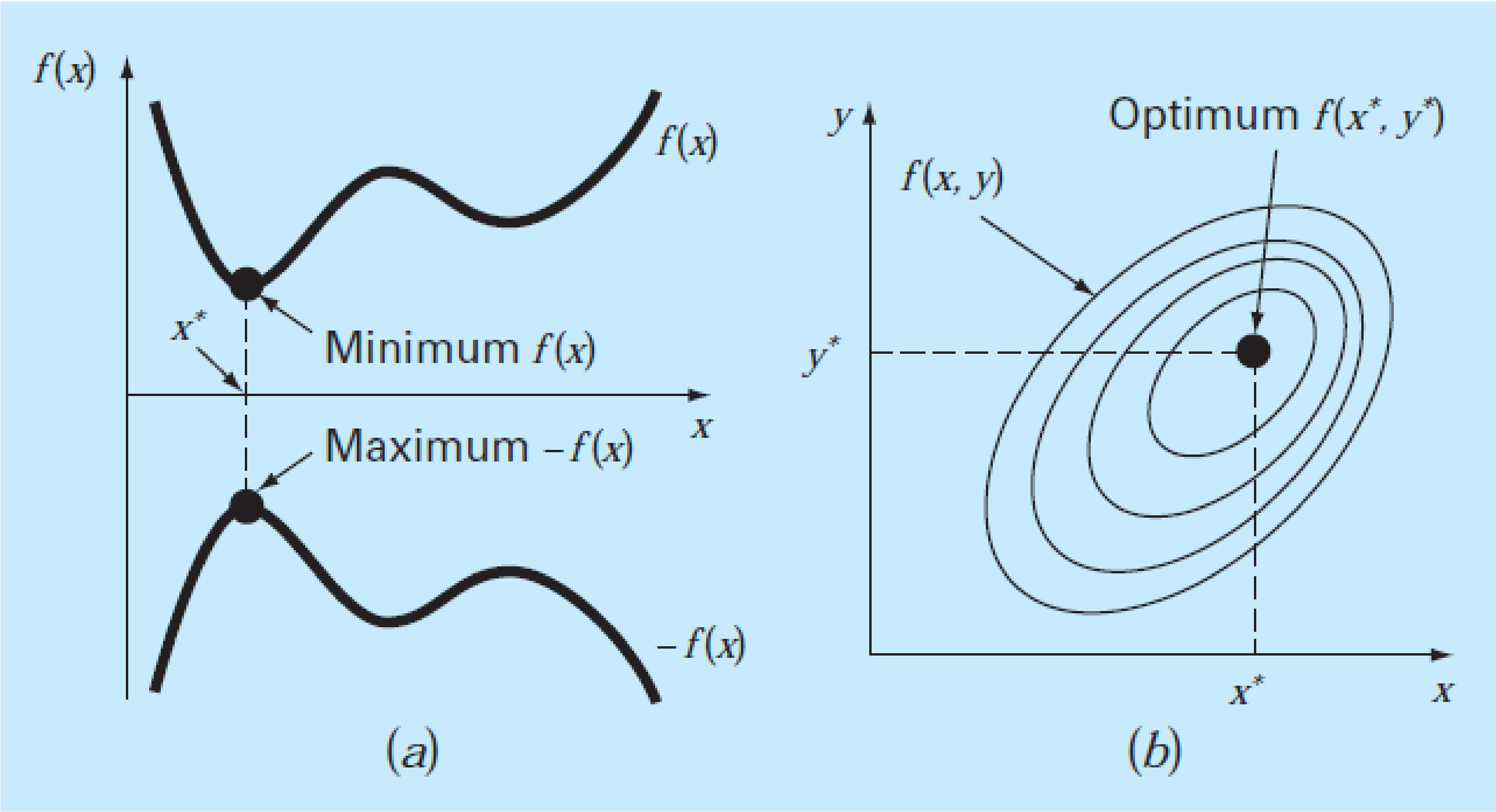
- Bir matris bir vektör ile çarpıldığında vektörün, davranışı; yönü, şiddeti, en önemlisi doğrultusu değişir.
- Bazı vektörler bir **A** matrisi ile çarpıldıkları zaman doğrultusu değişmez, yönü değişir. **Bu özel x vektörleri, Ax vektörü ile aynı yönde ya da ters yönde kalmaktadır. İşte bu vektörlere "özvektörler" denir.**
- Özdeğerler, bir matrisin orijinal yapısını görmek için kullanılan alternatif bir yoldur.
- Bir özvektörün **A** matrisi ile çarpımı olan A**x** vektörü, orijinal **x** vektörünün λ∊R olmak üzere λ katıdır.

Özdeğer ve özvektörün geometrik yorumu

$\underline{A}\ \underline{x}=\lambda\underline{x}$ bağıntısından hesaplan λ özdeğeri ve $\underline{x}$ özvektörü şu şekilde yorumlanabilir: $\underline{A}$ matrisi $\underline{x}$ vektörünü λ kadar büyütmekte veya küçültmektedir. $\underline{x}$ vektörünün doğrultusu değişmemekte fakat yönü değişebilmektedir. λ pozitif ise $\underline{x}$ ve $\lambda\underline{x}$ aynı yönde, aksi hale ters yöndedirler.



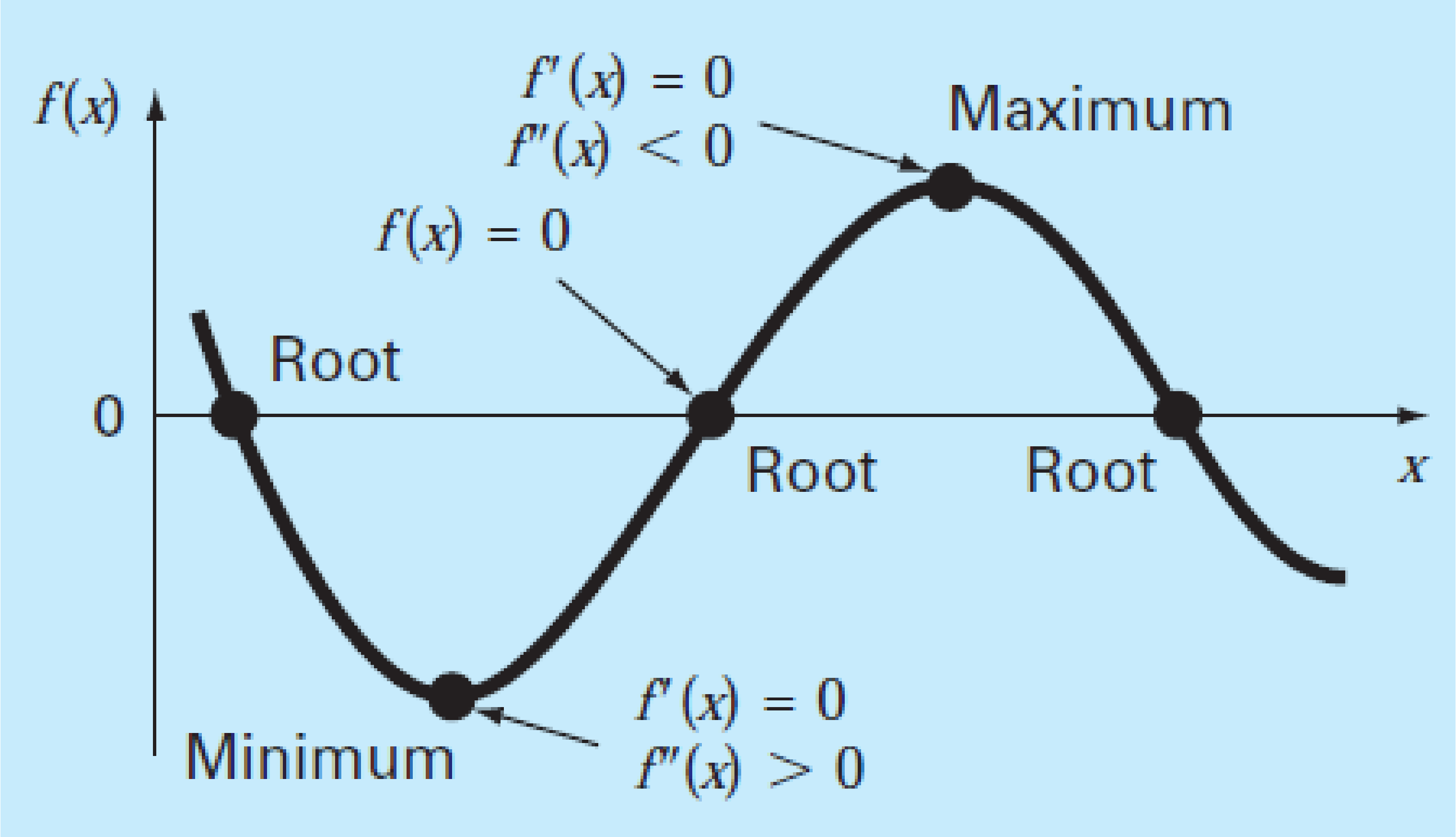
Tek Boyutlu Optimizasyon/Çok Boyutlu Optimizasyon



Matematiksel ifadelerde ya da denklemlerde
Minimum, maksimum, dönüm noktası,
denge noktası,
belirsizliklerin belirlenmesine
yönelik çalışmalar yapılır.

Optimizasyon: Matematiksel Tanım

- Her hangi bir matematiksel modelin grafiğini çizdiğinizde, maksimum ve minimum noktalarıyla karşılaşırız. Maksimum minimumların tepe noktası, türevin sıfır olduğu optimum noktalarıdır. Birinci türev f'(x)=0 optimum noktasını belirlerken f''(x)'de optimumun maksimum mu minimum mu olduğunu belirler.



Kök belirleme ve optimizasyon bir bakıma birbirine benzer, Kök belirlemede, eğrinin ekseni kestiği noktaları buluyorduk; Optimizasyon problemlerinde ise maximum veya minimum noktalarını bulacağız.

Optimizasyon: Matematiksel Tanım

- Bir optimizasyon problemi genel olarak şu şekilde ifade edilir.
- f(x)'i
  - $d_i(x) \leq a_i$ i=1,2,…,m
  - $e_{i(x)} = b_i$ i=1,2,…,p

şartları altında minimum ve maksimum kılacak x'i bulun, burada x n boyutlu bir tasarım vektörü, f(x) amaç fonksiyonu, $d_i(x)$'ler eşitsizlik şeklinde ifade edilen kısıtlar, $e_i(x)$'ler eşitlik şeklinde ifade edilen kısıtlar ve a_i ve b_i sabitlerdir. Fonksiyon için herhangi bir kısıt verilmezse bu tür optimizasyon problemlerine kısıtlamasız optimizasyon denir.

Kısıtlamalı problemlerde, serbestlik derecesi $n - m - p$ şeklinde bulunur.

Genellikle bir çözüm elde edebilmek için $m + p \leq n$ olmalıdır. $m + p \geq n$ olursa bu tür problemlere aşırı kısıtlı problem denir.

Newton Yöntemi

- Hatırla: Kök bulmada kullandığımız Newton-Raphson yöntemi

$$x_{i+1} = x_i - \frac{f(x)}{f'(x)}$$

- Yeni bir fonksiyon tanımlayalım $\mathrm{g(x)} = f'(x)$ olsun. Optimizasyon probleminde, $\mathrm{g}(x^*) = f'(x^*) = 0$'ı arıyoruz. Dolayısıyla,

$$x_{i+1} = x_i - \frac{f'(x)}{f''(x)}$$

yazabiliriz.

- Tek bir başlangıç tahmini yeterlidir.
- Yöntem hızlıdır, ancak ilk tahmin iyi değilse ıraksayabilir.
- Türev almak sıkıntı olursa, yaklaşık türev ifadeleri kullanılabilir.
- Iraksama problemlerini gidermek üzere hibrit yöntemler önce kapalı yöntemlerle optimum noktaya yaklaşıp ardından Newton yöntemiyle optimuma hızlıca ulaşmayı tercih ederler.

Örnek Problem

- Newton Yöntemi ve x_0=2.5 başlangıç tahminini kullanarak $f(x) = 2sinx - \frac{x^2}{10}$ fonksiyonunun maksimumunu bulun.

$$f'(x) = 2cosx - \frac{2x}{10} = 2cosx - \frac{x}{5}$$

$$f''(x) = -2sinx - \frac{2}{10} = -2sinx - \frac{1}{5}$$

i	x	f(x)	f'(x)	f''(x)
0	2,50000	0,57194	-2,10229	-1,39694
1	0,99508	1,57859	0,88985	-1,87761
2	1,46901	1,77385	-0,09058	-2,18965
3	1,42764	1,77573	-0,00020	-2,17954
4	1,42755	1,77573	0,00000	-2,17952

Türevin Yorumu

Birinci Türev

- Birinci ve ikinci türevlerinin verdiği bilgilerden f'(x) veya df/dx olarak yazılan f(x) fonksiyonunun ilk türevi, x noktasındaki teğet çizgisi fonksiyonun eğimidir.
- Grafik olmayan terimlerle ifade etmek gerekirse, ilk türev bize bir fonksiyonun nasıl arttığını veya azaldığını ve ne kadar artacağını veya azalacağını söyler.
- Pozitif eğim bize x arttıkça f(x)'nin de arttığını söyler. Negatif eğim bize x arttıkça f(x)'in azaldığını söyler. Sıfır eğim bize özel bir şey söylemez: fonksiyon o noktada artar, ne azalır veya yerel maksimumda veya yerel minimumda olabilir.

Türevler açısından bu bilgileri yazarken şunu görüyoruz:

-
- $\frac{df(p)}{dx} > 0$, ise f(x), x = p'de artan bir fonksiyondur.
- $\frac{df(p)}{dx} < 0$, ise f(x), x = p'de azalan bir fonksiyondur.
- $\frac{df(p)}{dx} = 0$, ise o zaman x = p, f (x)'in kritik noktası olarak adlandırılır ve x(p)'deki f (x) 'nin davranışı hakkında yorum yapılabilir. Eğer f''(X) = 0 ise, o zaman testin kesin olmadığını söyleyebiliriz. Şimdi, f''(X)'in işareti bu noktanın etrafında negatiften pozitife veya pozitiften negatife değişirse, noktanın bir dönüm noktası olması mümkündür. Fakat, f''(X)'in işareti değişmezse, bunun bir dönüm noktası olmadığını söyleyebiliriz.

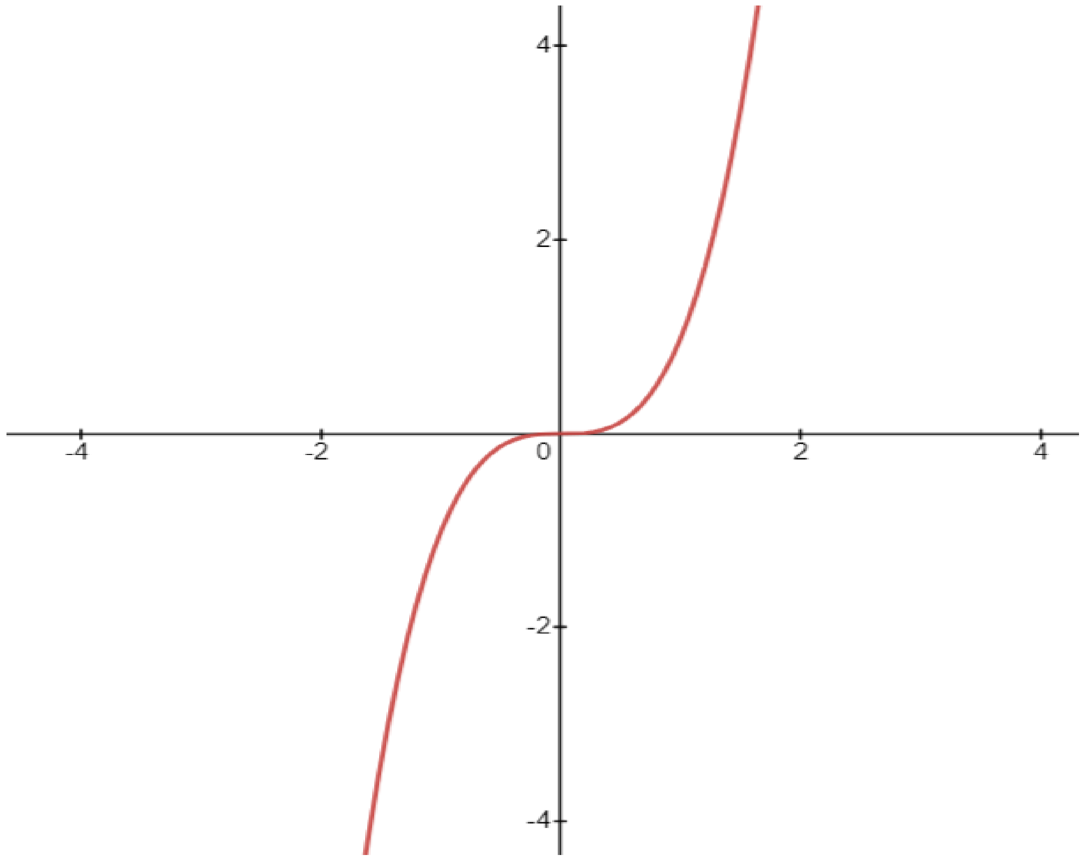
Türevin Yorumu

İkinci Türev

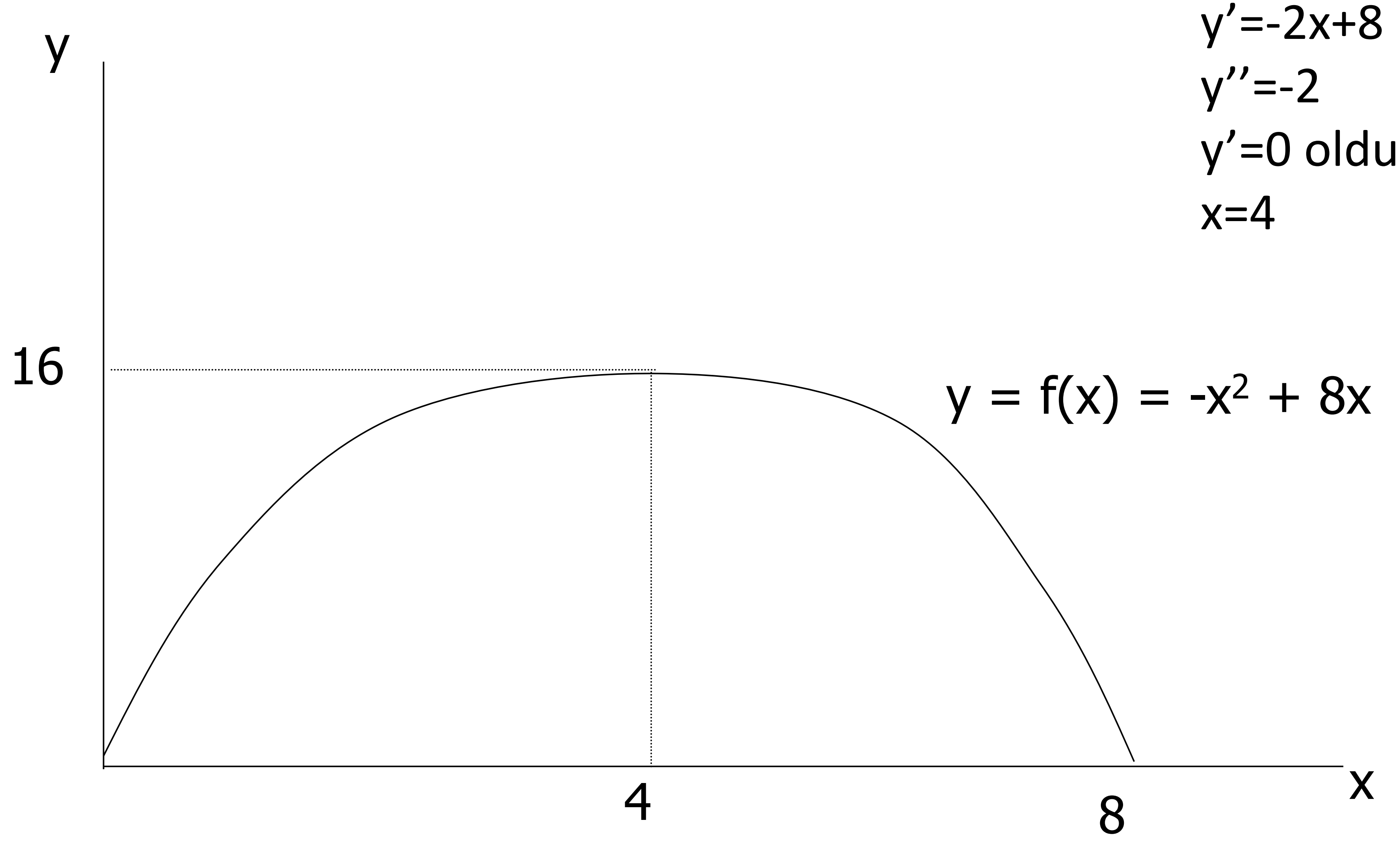
- Bir fonksiyonun ikinci türevi, f’’ (x) veya $\frac{d^2 f}{dx^2}$ olarak yazılır. İlk türev bize fonksiyonun arttığını veya azaldığını söylese de, ikinci türev,
- x = p'de $\frac{d^2 f(p)}{dx^2}$ >0 ise, f(x), x = p'de aşağı doğru kavislidir.
- x = p'de $\frac{d^2 f(p)}{dx^2}$ <0 ise, f(x), x = p'de yukarıdoğru kavislidir.
- x = p'de $\frac{d^2 f(p)}{dx^2}$ =0 ise, o zaman f(x) 'in x = p'deki davranışı hakkında bir yorum yapamıyoruz.
- Birinci türevin anlamından x, f(x) fonksiyonunun kritik bir noktası olduğunda, o noktada fonksiyonun davranışı hakkında bir yorum yapabilmek için, x’in bölgesel maksimum veya bölgesel minimum olduğunu öğrenmek için genellikle işlevin ikinci türevi kullanılır. Eğer f''(X)=0 ise ikinci türev testi kesin sonuç vermez. Dönüm noktalarını kontrol edin.

Dönüm Noktası

- Bir dönüm noktası, eğri üzerinde içbükeyliğin işaretinin değiştiği bir nokta olarak tanımlanır. Dönüm noktaları yerel maksimumlar veya yerel minimumlar olamaz. Aşağıdaki görüntüde, $f(X) = X^3$ fonksiyonu için $X = 0$'ın bir dönüm noktası olduğunu görebiliriz. Ayrıca, $(-\infty,0)$ aralığında fonksiyonun içbükey olduğunu ve $(0, \infty)$ aralığında fonksiyonun dışbükey olduğunu görebiliriz.



Graphical Representation of a Maximum



$y' = -2x + 8$

$y'' = -2$

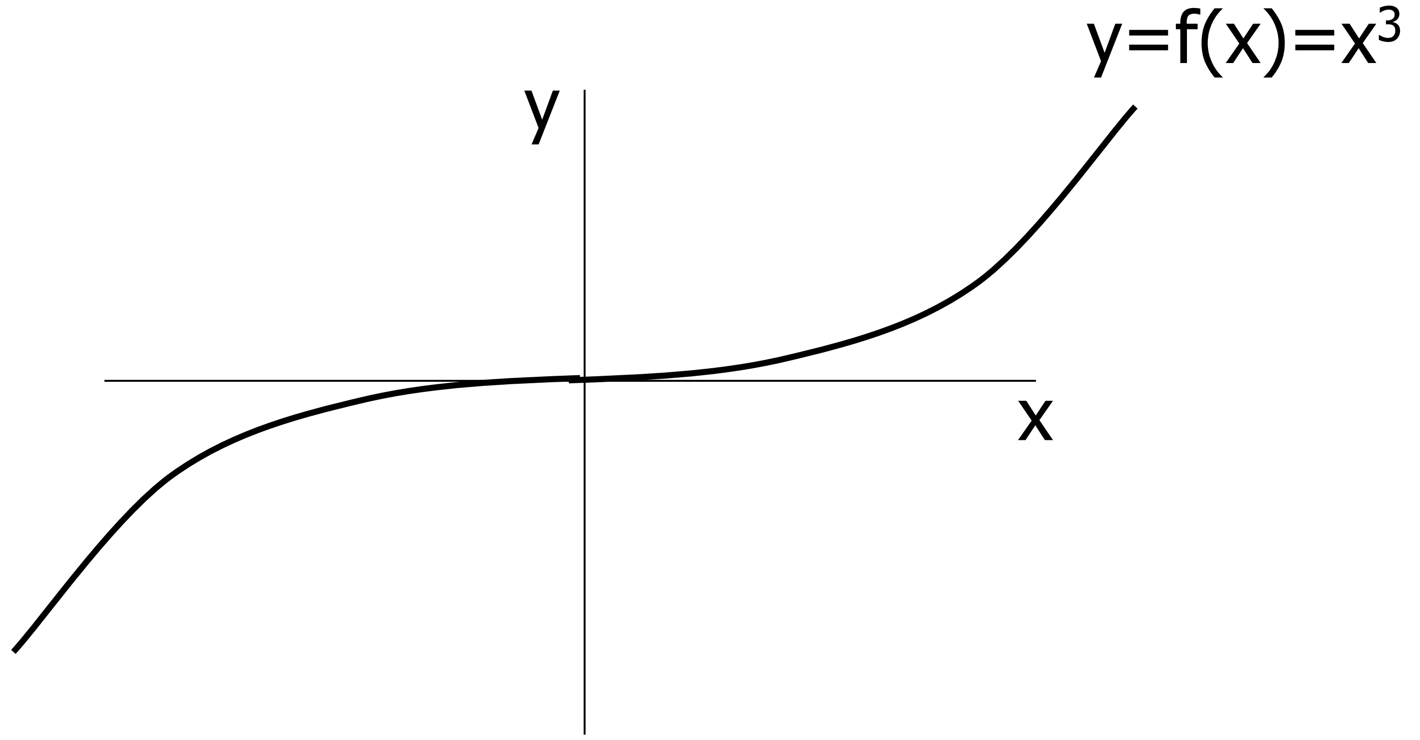
$y' = 0$ olduğu nokta $-2x + 8 = 0$

$x = 4$

$y = f(x) = -x^2 + 8x$

An Example of f''(x*)=0

- Suppose $y = f(x) = x^3$, then $f'(x) = 3x^2$ and $f''(x) = 6x$,
  - This implies that $x^* = 0$ and $f''(x^*=0) = 0$.

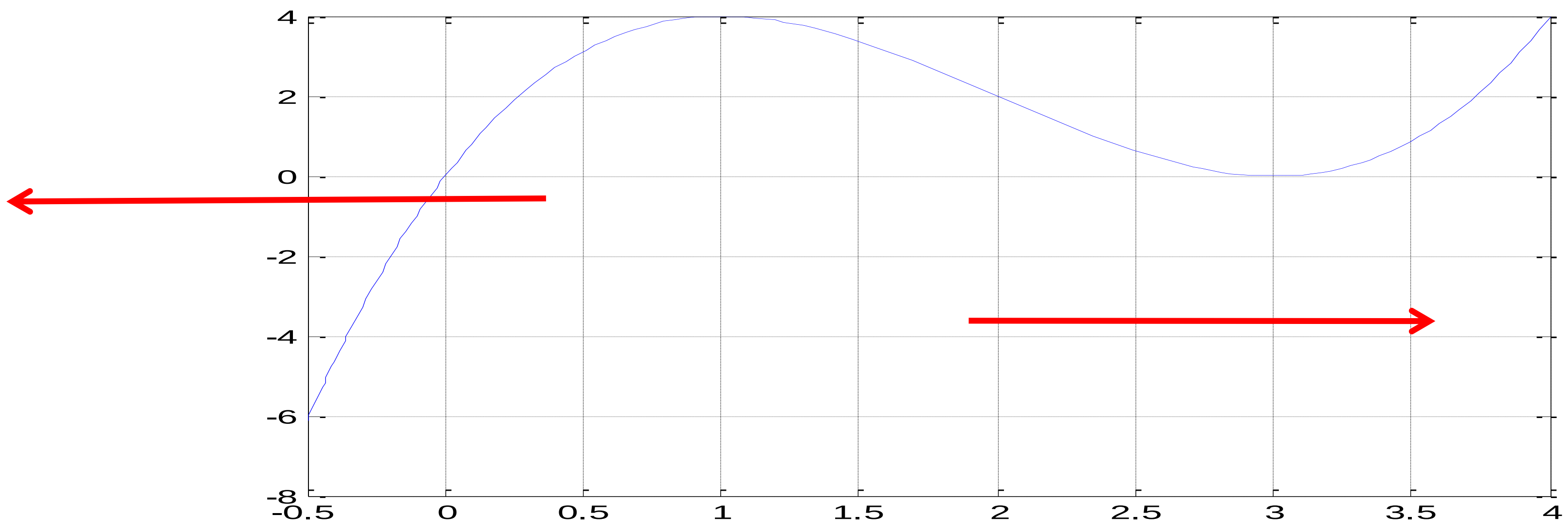


x*=0 is a saddle point where the point is neither a maximum nor a minimum

Example of Using First and Second Order Conditions

- Suppose you have the following function:
  - $f(x) = x^3 – 6x^2 + 9x$
- Then the first order condition to find the critical points is:
  - $f'(x) = 3x^2 - 12x + 9 = 0$
  - This implies that the critical points are at x = 1 and x = 3.

Amacımız, birinci ve ikinci türevlere bakarak fonksiyonun davranış hakkında yorum yapmaktır.



Example of Using First and Second Order Conditions (Cont.)

- The next step is to determine whether the critical points are maximums or minimums.
  - These can be found by using the second order condition.
  - $f''(x) = 6x – 12 = 6(x-2)$

- Testing x = 1 implies:
  - $f''(1) = 6(1-2) = -6 < 0$.
  - Hence at x =1, we have a maximum.
- Testing x = 3 implies:
  - $f''(3) = 6(3-2) = 6 > 0$.
  - Hence at x =3, we have a minimum.

- Are these the ultimate maximum and minimum of the function f(x)?

Conditions for a Minimum or a Maximum Value of a Function of Several Variables

- Correspondingly, for a function $f(\mathbf{x})$ of several independent variables $\mathbf{x}$
  - Calculate $\nabla f(\mathbf{x})$ and set it to zero.
  - Solve the equation set to get a solution vector $\mathbf{x}^*$.
  - Calculate $\nabla^2 f(\mathbf{x})$.
  - Evaluate it at $\mathbf{x}^*$.
  - Inspect the Hessian matrix at point $\mathbf{x}^*$.

$$\mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x})$$

Hessian Matrix of f(x)

$f(\mathbf{x})$ is a C^2 function of n variables,

$$\mathbf{H}(\mathbf{x}) \equiv \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}.$$

Since cross-partials are equal for a C^2 function, H(x) is a symmetric matrix.

Conditions for a Minimum or a Maximum Value of a Function of Several Variables (cont.)

- Let f(x) be a C^2 function in R^n. Suppose that x* is a critical point of f(x), i.e., $\nabla f(\mathbf{x}^*) = 0$

1. If the Hessian $\mathbf{H}(\mathbf{x}^*)$ is a positive definite matrix, then **x*** is a local minimum of f(x);
2. If the Hessian $\mathbf{H}(\mathbf{x}^*)$ is a negative definite matrix, then **x*** is a local maximum of f(x).
3. If the Hessian $\mathbf{H}(\mathbf{x}^*)$ is an indefinite matrix, then **x*** is neither a local maximum nor a local minimum of f(x).

Example

- Find the local maxs and mins of f(x,y)

$$f(x, y) = x^3 - y^3 + 9xy$$

- Firstly, computing the first order partial derivatives (i.e., gradient of f(x,y)) and setting them to zero

$$\nabla f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 3x^2 + 9y \\ -3y^2 + 9x \end{pmatrix} = 0$$

$$\Rightarrow \text{critical points } (x^*, y^*) \text{ is } (0,0) \text{ and } (3, -3).$$

Example (Cont.)

- We now compute the Hessian of f(x,y)

$$\nabla^2 f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}.$$

- The first order leading principal minor is 6x and the second order principal minor is -36xy-81.
- At (0,0), these two minors are 0 and -81, respectively. Since the second order leading principal minor is negative, (0,0) is a saddle of f(x,y), i.e., neither a max nor a min.
- At (3, -3), these two minors are 18 and 243. So, the Hessian is positive definite and (3,-3) is a local min of f(x,y).
- Is (3, -3) a global min?

Global Maxima and Minima of a Function of Several Variables

- Let f(x) be a C^2 function in R^n, then

- When f(x) is a concave function, i.e., $\nabla^2 f(\mathbf{x})$ is negative semidefinite for all x and $\nabla f(\mathbf{x}^*) = 0$, then x* is a global max of f(x);

- When f(x) is a convex function, i.e., $\nabla^2 f(\mathbf{x})$ is positive semidefinite for all x and $\nabla f(\mathbf{x}^*) = 0$, then x* is a global min of f(x);

Example (Discriminating Monopolist)

- A monopolist producing a single output has two types of customers. If it produces q_1 units for type 1, then these customers are willing to pay a price of $50-5q_1$ per unit. If it produces q_2 units for type 2, then these customers are willing to pay a price of $100-10q_2$ per unit.
- The monopolist's cost of manufacturing q units of output is 90+20q.
- In order to maximize profits, how much should the monopolist produce for each market?
- Profit is:

$$f(q_1,q_2)=q_1(50-5q_1)+q_2(100-10q_2)-(90+20(q_1+q_2)).$$

The critical points are

$$\frac{\partial f}{\partial q_1}=50-10q_1-20=0 \Rightarrow q_1=3,\quad \frac{\partial f}{\partial q_2}=100-20q_2-20=0 \Rightarrow q_2=4.$$

$$\frac{\partial^2 f}{\partial q_1^2}=-10,\quad \frac{\partial^2 f}{\partial q_2^2}=-20,\quad \frac{\partial^2 f}{\partial q_1 \partial q_2}=\frac{\partial^2 f}{\partial q_2 \partial q_1}=0.$$

$$\Rightarrow \nabla^2 f \text{ is negative definite} \Rightarrow (3,4) \text{ is the profit-maximizing supply plan.}$$

Kısıtlı Optimizasyon

Constrained Optimization

- **Examples:**
- Individuals maximizing utility will be subject to a budget constraint
- Firms maximising output will be subject to a cost constraint

- The function we want to maximize/minimize is called the objective function
- The restriction is called the constraint

Constrained Optimization (General Form)

- A general *mixed constrained* multi-dimensional maximization problem is

$$
\begin{aligned}
&\max \; f(x) = f(x_1, \ldots, x_n) \\
&\text{subject to} \\
&g_1(x_1, \ldots, x_n) \le b_1,\; g_2(x_1, \ldots, x_n) \le b_2, \cdots, g_k(x_1, \ldots, x_n) \le b_k, \\
&h_1(x_1, \ldots, x_n) = c_1,\; h_2(x_1, \ldots, x_n) = c_2, \cdots, h_m(x_1, \ldots, x_n) = c_m.
\end{aligned}
$$

Constrained Optimization (Lagrangian Form)

- The Lagrangian approach is to associate a *Lagrange multiplier* λ_i with the i^{th} inequality constraint and μ_i with the i^{th} equality constraint.
- We then form the *Lagrangian*

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_m) =$$

$$f(x_1, \ldots, x_n) - \sum_{i=1}^{k} \lambda_i \left(g_i(x_1, \ldots, x_n) - b_i \right)$$

$$- \sum_{i=1}^{m} \mu_i \left(h_i(x_1, \ldots, x_n) - c_i \right).$$

Constrained Optimization (Kuhn-Tucker Conditions)

- If x^* is a local maximum of f on the constraint set defined by the k inequalities and m equalities, then, there exists multipliers $\lambda_1^*, \cdots, \lambda_k^*, \mu_1^*, \cdots, \mu_m^*$ satisfying

$$\frac{\partial L(x^*, \lambda^*, \mu^*)}{\partial x_j} = \frac{\partial f(x^*)}{\partial x_j} - \sum_{i=1}^{k} \lambda_i^* \frac{\partial g_i(x^*)}{\partial x_j} - \sum_{i=1}^{m} \mu_i^* \frac{\partial h_i(x^*)}{\partial x_j} = 0, \quad j = 1, \ldots, \mathrm{n}$$

$$h_i(x^*) = c_i, \quad i = 1, \ldots, m$$

$$\lambda_i^* \left(g_i(x^*) - b_i \right) = 0, \quad i = 1, \ldots, k$$

$$g_i(x^*) \le b_i, \quad i = 1, \ldots, k$$

$$\lambda_i^* \ge 0, \quad i = 1, \ldots, k$$

29

Constrained Optimization (Kuhn-Tucker Conditions)

- The first set of KT conditions generalizes the unconstrained critical point condition
- The second set of KT conditions says that x needs to satisfy the equality constraints
- The third set of KT conditions is

$$\lambda_i^*\left(g_i(x^*)-b_i\right)=0, \quad i=1,\ldots,k$$

- That is to say

$$\text{if } \lambda_i^* > 0 \text{ then } g_i(x^*) = b_i$$

$$\text{if } g_i(x^*) < b_i \text{ then } \lambda_i^* = 0$$

Constrained Optimization (Kuhn-Tucker Conditions)

- This can be interpreted as follows:
- Additional units of the *resource* b_i only have value if the available units are used fully in the optimal solution, i.e., if

$g_i(x^*) < b_i$ the constraint is not binding thus it does not make difference in the optimal solution and λ_i*=0.

- Finally, note that increasing b_i enlarges the feasible region, and therefore increases the objective value
  - Therefore, $\lambda_i \geq 0$ for all i

Example

$$\max \ x - y^2$$
subject to
$$x^2 + y^2 = 4$$
$$x \geq 0, y \geq 0.$$

Form the Lagrangian

L=x-y^2 $- \mu(x^2 + y^2 - 4) + \lambda_1 x + \lambda_2 y.$

The first order conditions become:

$$(1) \frac{\partial L}{\partial x} = 1 - 2\mu x + \lambda_1 = 0,$$

$$(2) \frac{\partial L}{\partial y} = -2y - 2\mu y + \lambda_2 = 0,$$

$$(3) x^2 + y^2 - 4 = 0$$

$(4) \lambda_1 x = 0,$ $(5) \lambda_2 y = 0,$

$(6) \lambda_1 \geq 0,$ $(7) \lambda_2 \geq 0,$

$(8) x \geq 0,$ $(9) y \geq 0.$

32

Example (cont.)

By (1), $1+\lambda_1 = 2\mu x$, since $\lambda_1 \geq 0, 1+\lambda_1 > 0$

$\Rightarrow \mu > 0$ and $x > 0$, by (4), $\lambda_1 = 0$.

by (2), $\lambda_2 = 2y(1+\mu)$, since $1+\mu > 0$

$\Rightarrow$ either both y and λ_2 are zero, or both

are positive, from (5) $\Rightarrow y = 0, \lambda_2 = 0$.

By (3) and (8) $\Rightarrow$ x=2, from (4), $\lambda_1 = 0$,

by (1), $\mu = \dfrac{1}{4}$. So, $(x, y, \mu, \lambda_1, \lambda_2) = (2, 0, \dfrac{1}{4}, 0, 0)$.

Sensitivity Analysis

Sensitivity Analysis

- We notice that

$$L(x^*, \lambda^*, \mu^*) = f(x^*) - \lambda^{*\prime}\left(g(x^*) - b\right) - \mu^{*\prime}(h(x^*) - c) = f(x^*)$$

- What happens to the optimal solution value if the right-hand side of constraint *i* is changed by a *small* amount, say Δb_i, *i=1⋯k.* *or* Δc_i, *i=1⋯m.*
  - It changes by *approximately* $\lambda_i^* \Delta b_i$ *or* $\mu_i^* \Delta c_i$
  - λ_i^* is the *shadow price* of i^th inequality constraint *and* μ_i^* is the *shadow price* of i^th equality constraint

Sensitivity Analysis (Example)

- In the previous example, if we change the first constraint to $x^2+y^2=3.9$, then we predict that the new optimal value would be 2+1/4(-0.1)=1.975.
- If we compute that problem with this new constraint, then $x-y^2= \sqrt{3.9} = 1.9748.$
- If, instead, we change the second constraint from $x\geq 0$ to $x\geq 0.1$, we do not change the solution or the optimum value since $\lambda_1^* = 0.$

Utility Maximization Example

The utility derived from exercise (X) and watching movies (M) is described by the function

$$U(X,M) = 100 - e^{-2X} - e^{-M}$$

Four hours per day are available to watch movies and exercise.

Our Lagrangian function is

$$L(X,M,\lambda) = 100 - e^{-2X} - e^{-M} - \lambda(X + M - 4)$$

First-Order Conditions:

$$L_X = 2e^{-2X} - \lambda = 0$$

$$L_M = e^{-M} - \lambda = 0$$

$$L_\lambda = X + M - 4 = 0$$

Utility Max Example Continued

First - Order Conditions :

$L_X = 2e^{-2X} - \lambda = 0,\ (1) \quad L_M = e^{-M} - \lambda = 0,\ (2) \quad L_\lambda = X + M - 4 = 0\ (3)$

From (2) we get that $\lambda = e^{-M}$. Substituting into (1), we get

$2e^{-2X} - e^{-M} = 0$ Solving (3) for M and substituting, we get

$2\mathrm{e}^{-2\mathrm{X}} - e^{-(4-X)} = 0 \rightarrow \ln(2) - 2\mathrm{X} = -4 + \mathrm{X}$ or $3\mathrm{X} = \ln(2) + 4$

$$\mathrm{X}^* = \frac{\ln(2)+4}{3} \textit{ and } M^* = \frac{8-\ln(2)}{3}$$

Sensitivity Analysis

Static optimisation: variables have numerical values, fixed with respect to time.

Dynamic optimisation: variables are functions of time.

Essential Features: Every optimisation problem contains three essential categories:

1. At least one objective function to be optimised
2. Equality constraints
3. Inequality constraints

Mathematical Description

Minimize : $f(\mathbf{x})$ objective function

Subject to: $\begin{cases} \mathbf{h}(\mathbf{x}) = \mathbf{0} & \text{equality constraints} \\ \mathbf{g}(\mathbf{x}) \geq \mathbf{0} & \text{inequality constraints} \end{cases}$

where $\mathbf{x} \in \Re^n$, is a vector of n variables $(x_1, x_2, \cdots, x_n)$

$\mathbf{h}(\mathbf{x})$ is a vector of equalities of dimension m_1

$\mathbf{g}(\mathbf{x})$ is a vector of inequalities of dimension m_2

Steps Used To Solve Optimisation Problems

1. Tüm değişkenlerin bir listesini yapmak için süreç analiz edilir.
2. Optimizasyon kriteri ve amaç fonksiyonu belirlenir.
3. Eşitlik ve eşitsizlik kısıtlamalarını tanımlamak için sürecin matematiksel modeli geliştirilir.
4. Serbestlik derecesi sayısını elde etmek için bağımsız ve bağımlı değişkenler belirlenir.
5. Problem formülasyonu çok büyük veya karmaşıksa, mümkünse basitleştirilir.
6. Uygun bir optimizasyon tekniği uygulanır.
7. Sonuç kontrol edilir ve model parametreleri ve varsayımlarındaki değişikliklere duyarlılığı incelenir.

Classification of Optimisation Problems

Properties of $f(\mathbf{x})$

- single variable or multivariable
- linear or nonlinear
- sum of squares
- quadratic
- smooth or non-smooth
- sparsity

Properties of h(x) and g(x)

- simple bounds
- smooth or non-smooth
- sparsity
- linear or nonlinear
- no constraints

Properties of variables x

- time variant or invariant
- continuous or discrete
- take only integer values
- mixed

Obstacles and Difficulties

- Objective function and/or the constraint functions may have finite *discontinuities* in the continuous parameter values.
- Objective function and/or the constraint functions may be *non-linear functions* of the variables.
- Objective function and/or the constraint functions may be defined in terms of *complicated interactions* of the variables. This may prevent calculation of unique values of the variables at the optimum.

- Objective function and/or the constraint functions may exhibit nearly “flat” behaviour for some ranges of variables or exponential behaviour for other ranges. This causes the problem to be insensitive, or too sensitive.
- The problem may exhibit many local optima whereas the global optimum is sought. A solution may be obtained that is less satisfactory than another solution elsewhere.
- Absence of a feasible region.
- Model-reality differences.

Typical Examples of Application static optimisation

- Plant design (sizing and layout).
- Operation (best steady-state operating condition).
- Parameter estimation (model fitting).
- Allocation of resources.
- Choice of controller parameters (e.g. gains, time constants) to minimise a given performance index (e.g. overshoot, settling time, integral of error squared).

dynamic optimisation

- Determination of a control signal $u(t)$ to transfer a dynamic system from an initial state to a desired final state to satisfy a given performance index.
- Optimal plant start-up and/or shut down.
- Minimum time problems

BASIC PRINCIPLES OF STATIC OPTIMISATION THEORY

Continuity of Functions

Functions containing discontinuities can cause difficulty in solving optimisation problems.

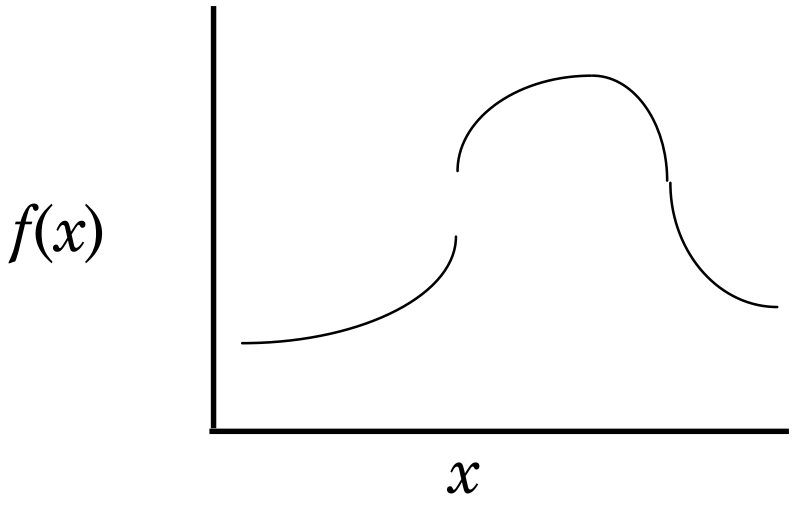
Definition: A function of a single variable x is continuous at a point x_o if:

$$\text{(a)} \quad f(x_o) \quad \text{exists}$$

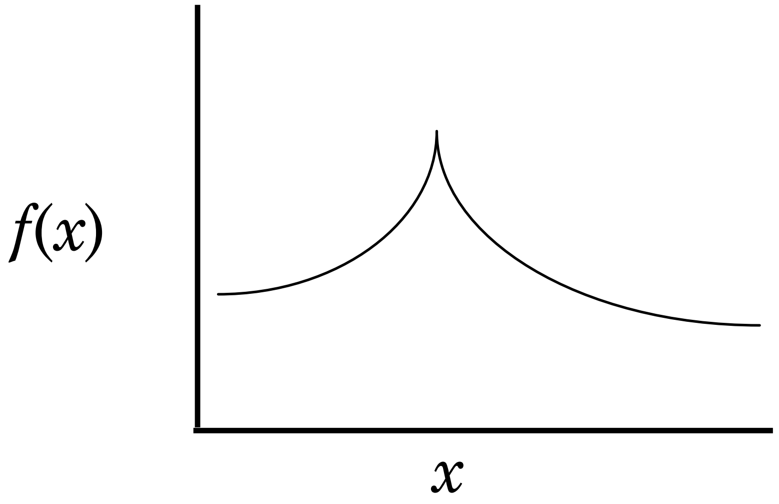
$$\text{(b)} \quad \lim_{x \to x_o} f(x) \quad \text{exists}$$

$$\text{(c)} \quad \lim_{x \to x_o} f(x) = f(x_o)$$

If $f(x)$ is continuous at every point in a region R, then $f(x)$ is said to be continuous throughout R.



$f(x)$ is discontinuous.



$f(x)$ is continuous, but $f'(x) \equiv \frac{df}{dx}(x)$ is not.

Unimodal and Multimodal Functions

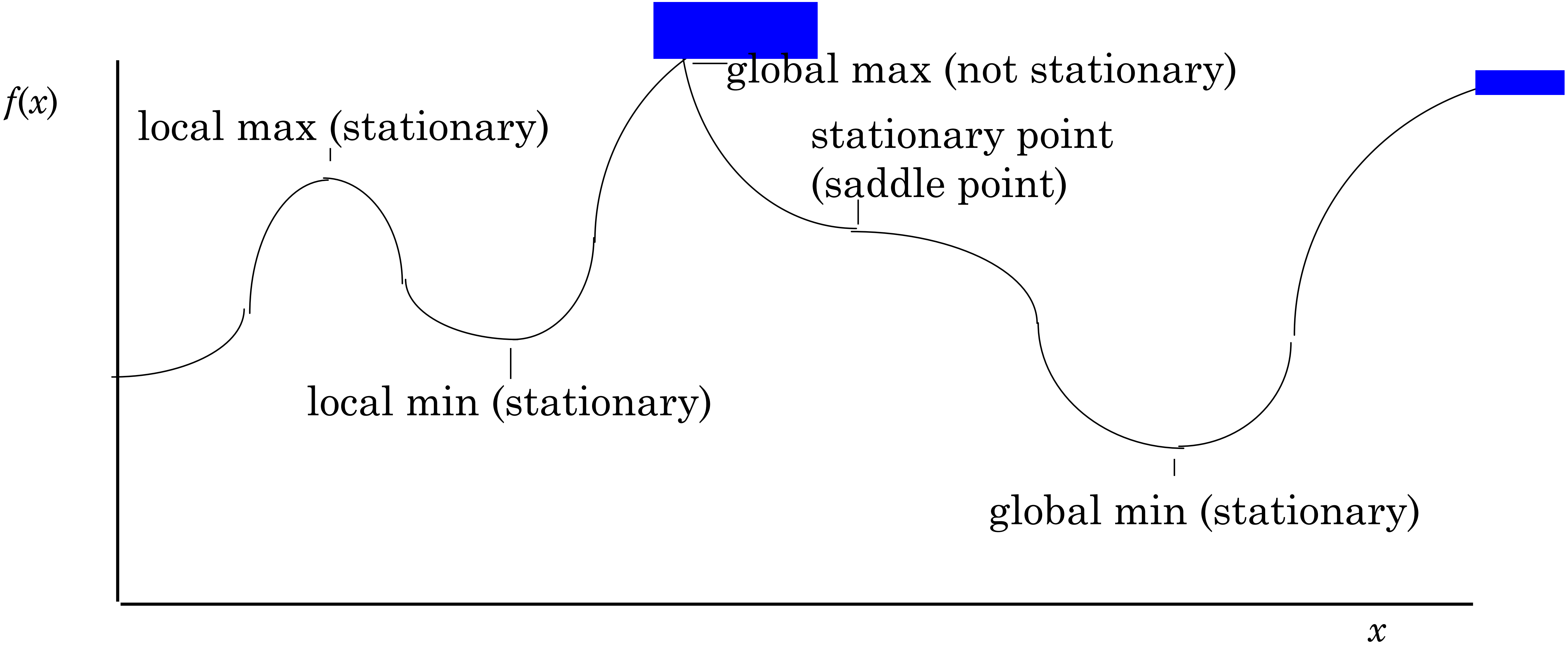
A *unimodal function* $f(x)$ (in the range specified for x) has a single extremum (minimum or maximum).

A *multimodal function* $f(x)$ has two or more extrema.

If $f'(x) = 0$ at the extremum, the point is called a *stationary point*.

There is a distinction between the *global extremum* (the biggest or smallest between a set of extrema) and *local extrema* (any extremum). *Note*: many numerical
procedures terminate at a local extremum.

A multimodal function



Multivariate Functions - Surface and Contour Plots

We shall be concerned with basic properties of a scalar function $f(\mathbf{x})$ of n variables $(x_1,\ldots,x_n)$.

If $n = 1$, $f(x)$ is a *univariate function*

If $n > 1$, $f(\mathbf{x})$ is a *multivariate function*.

For any multivariate function, the equation
$z = f(\mathbf{x})$ defines a *surface* in n+1 dimensional space .

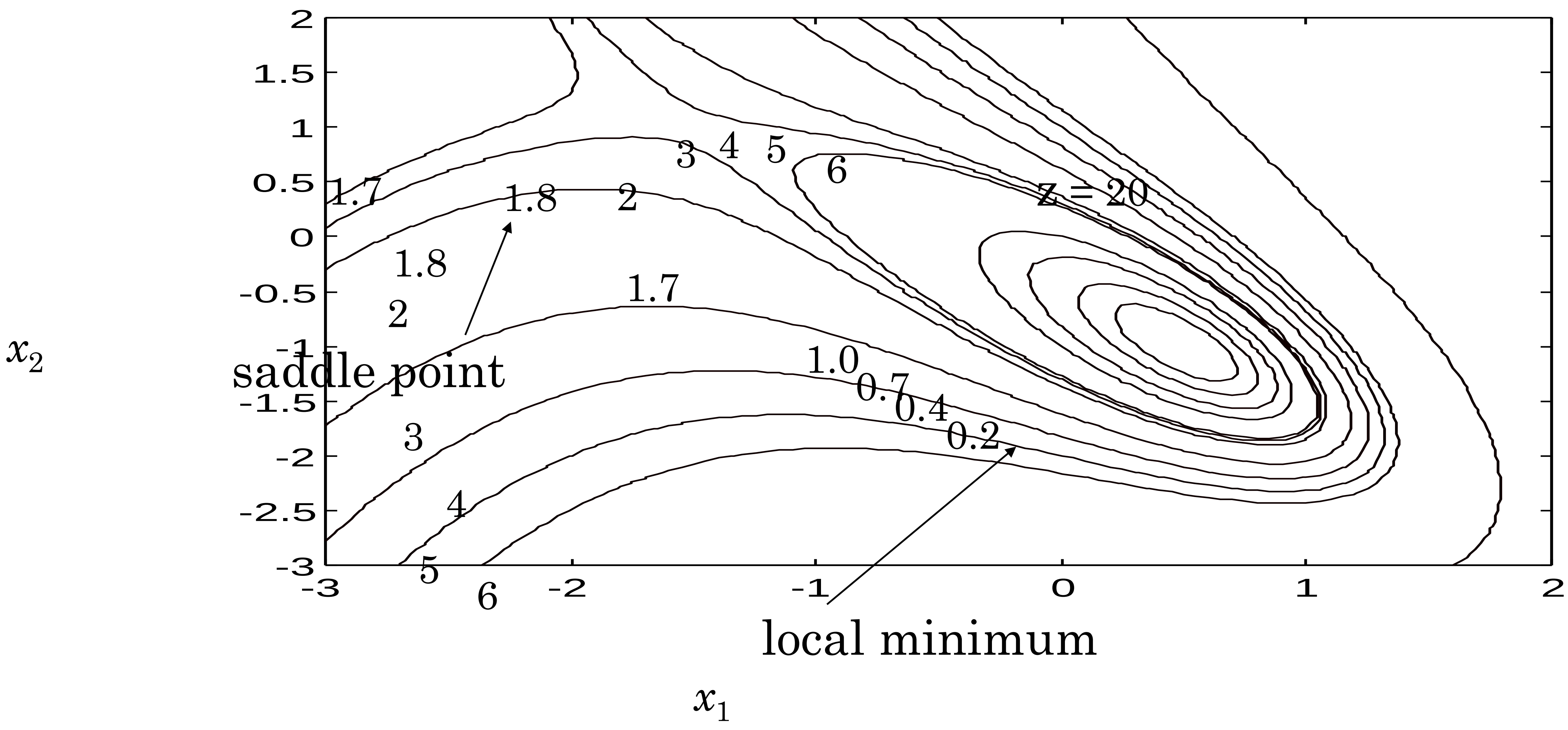
$$\Re^{n+1}$$

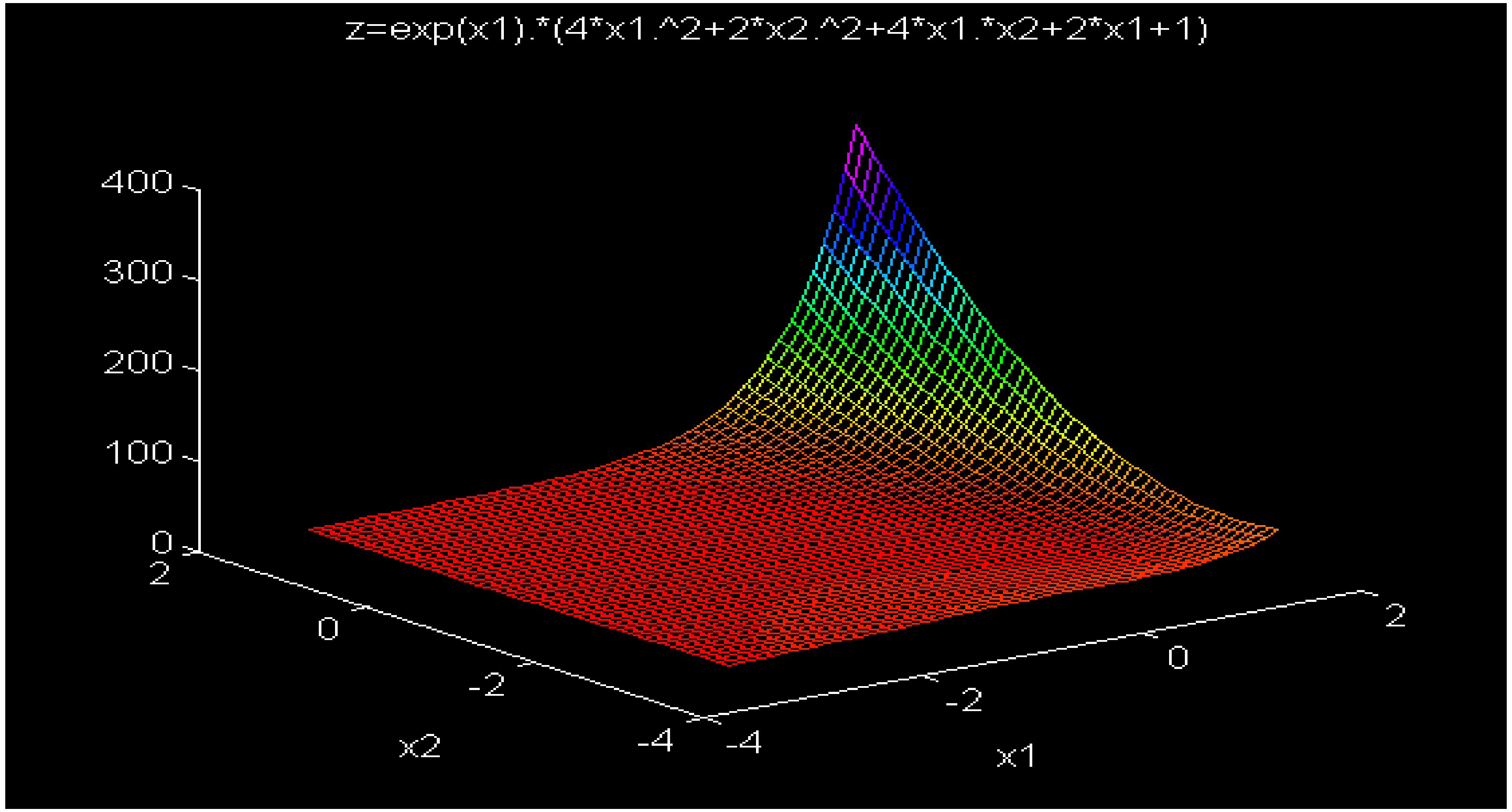
In the case $n = 2$, the points $z = f(x_1,x_2)$ represent a three dimensional surface.

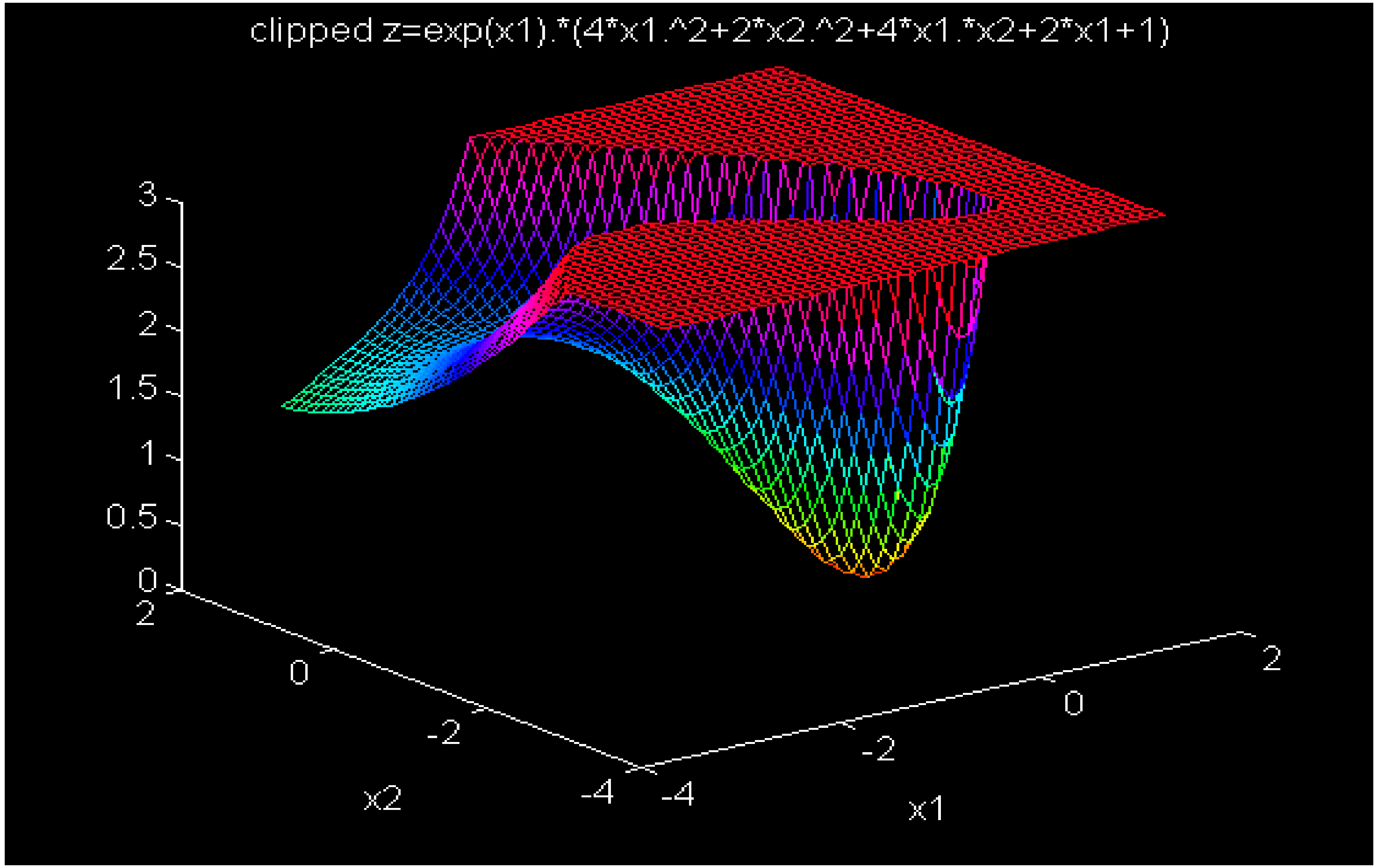
Let c be a particular value of $f(x_1,x_2)$. Then $f(x_1,x_2) = c$ defines a curve in x_1 and x_2 on the plane $z = c$.

If we consider a selection of different values of c, we obtain a family of curves which provide a *contour map* of the function $z = f(x_1,x_2)$.

contour map of $z = e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1)$

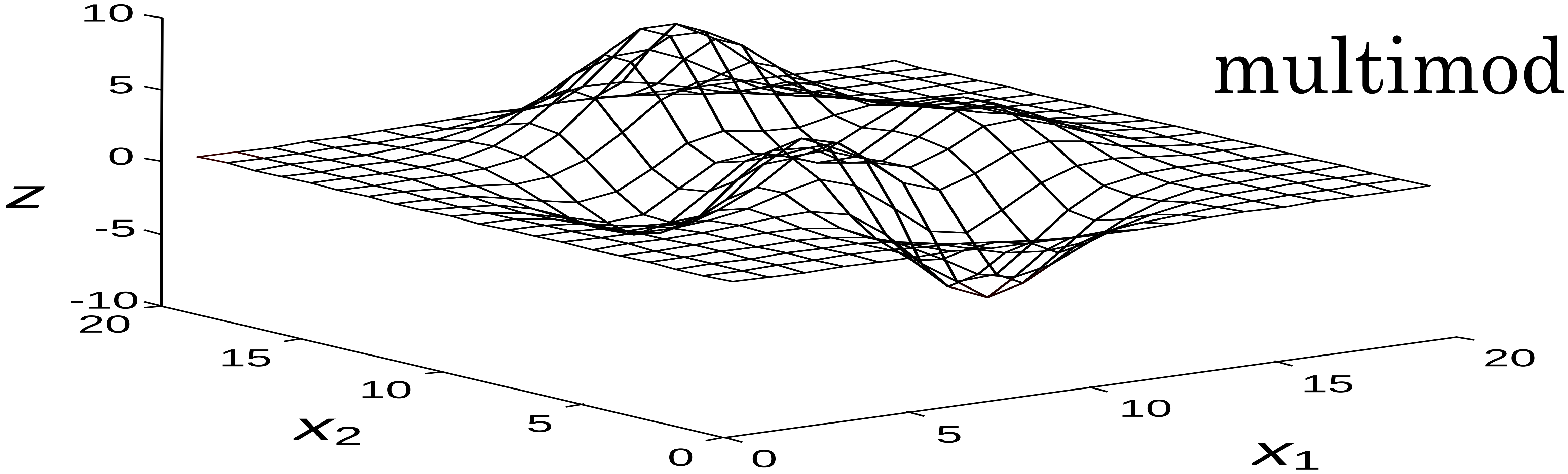



z=exp(x1).*(4*x1.^2+2*x2.^2+4*x1.*x2+2*x1+1)
400
300
200
100
0
2
0
-2
-4
x2
-4
-2
0
2
x1

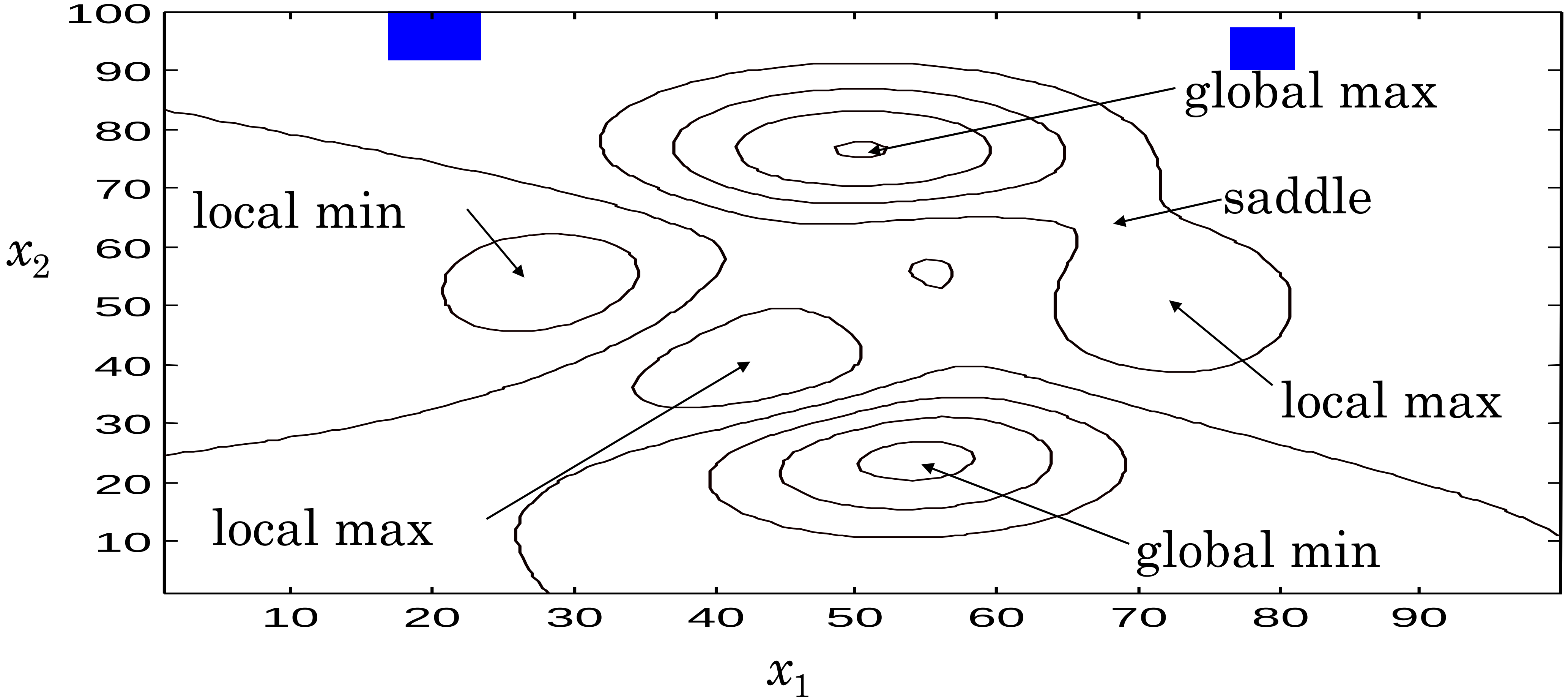

clipped z=exp(x1).*(4*x1.^2+2*x2.^2+4*x1.*x2+2*x1+1)
3
2.5
2
1.5
1
0.5
0
2
0
-2
-4
x2
-4
-2
0
2
x1

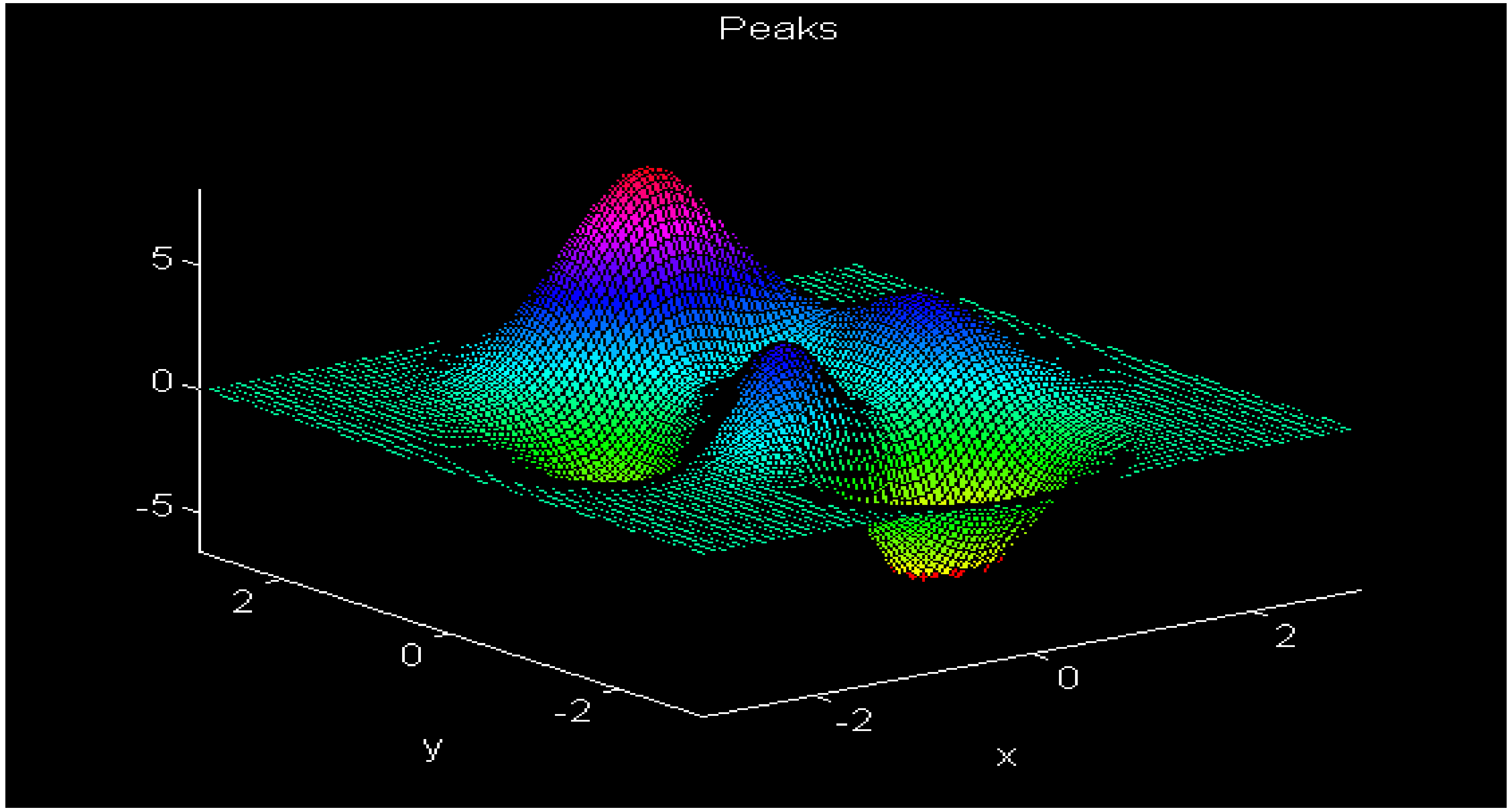
Example: Surface and Contour Plots of "Peaks" Function

$$z = 3(1 - x_1)^2 \exp\left(-x_1^2 - (x_2 + 1)^2\right)$$
$$- 10(0.2x_1 - x_1^3 - x_2^5)\exp(-x_1^2 - x_2^2)$$
$$- 1/3 \exp\left(-(x_1 + 1)^2 - x_2^2\right)$$


10
5
0
-5
-10
Z
20
15
10
5
0
x2
0
5
10
15
20
x1

multimodal!


100
90
80
70
60
50
40
30
20
10
x_2
10
20
30
40
50
60
70
80
90
x_1
global max
saddle
local min
local max
local max
global min


Peaks
5
0
-5
2
0
-2
y
-2
0
2
x

Gradient Vector

The slope of $f(\mathbf{x})$ at a point $\mathbf{x} = \bar{\mathbf{x}}$ in the direction of the i^{th} co-ordinate axis is

$$\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x} = \bar{\mathbf{x}}}$$

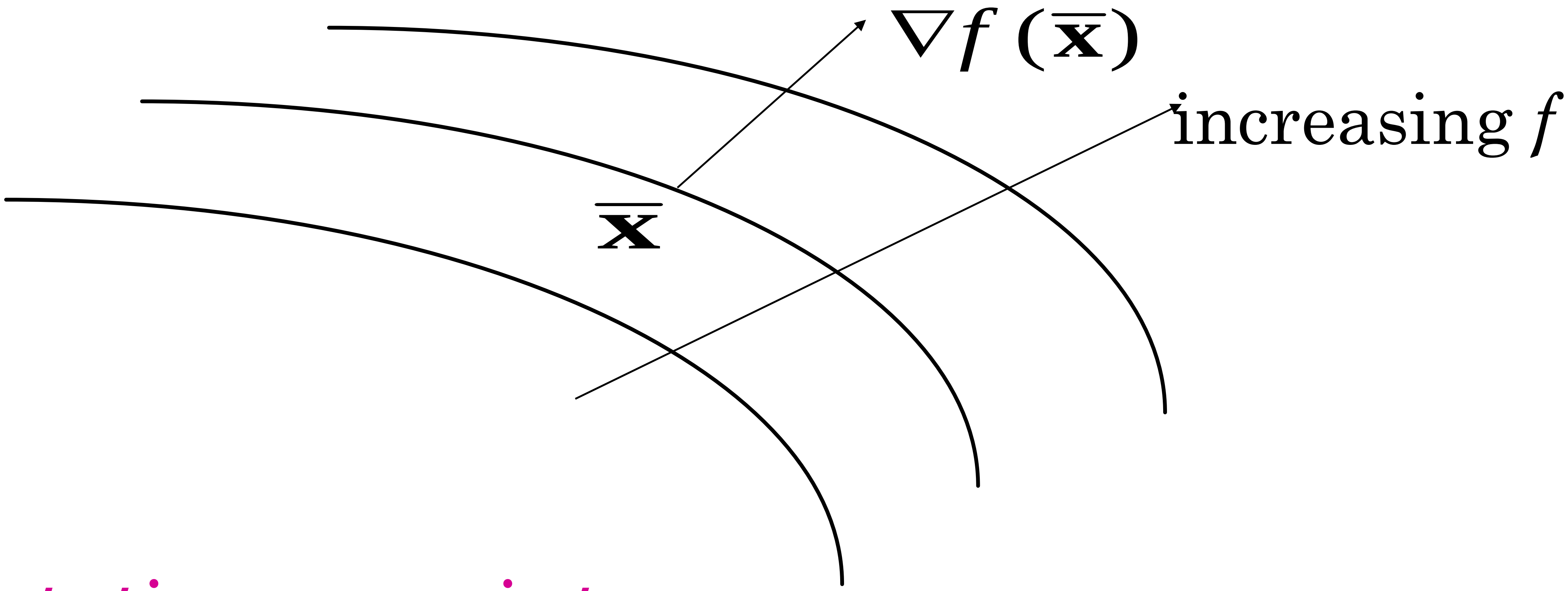
The n-vector of these partial derivatives is termed the *gradient vector* of f, denoted by:

$$\nabla f(\mathbf{x}) \equiv \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

(*a column vector*)

The gradient vector at a point $\mathbf{x} = \overline{\mathbf{x}}$

is normal to the the contour through that point in the direction of increasing *f*.



At a *stationary point*:

$$\nabla f(\mathbf{x}) = 0 \quad (\textit{a null vector})$$

Example

$$f(\mathbf{x}) = x_1 x_2^2 + x_2 \cos x_1$$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f(\mathbf{x})}{\partial x_1} \\ \dfrac{\partial f(\mathbf{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} x_2^2 - x_2 \sin x_1 \\ 2x_1 x_2 + \cos x_1 \end{bmatrix}$$

and the stationary point (points) are given by the simultaneous solution(s) of:

$$x_2^2 - x_2 \sin x_1 = 0$$

$$2x_1 x_2 + \cos x_1 = 0$$

Note: If $\nabla f(\mathbf{x})$ is a constant vector,

$f(\mathbf{x})$ is then *linear*.

e.g.

$$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \alpha \quad \Rightarrow \quad \nabla f(\mathbf{x}) = \mathbf{c}$$

Hessian Matrix (Curvature Matrix

The *second derivative* of a n - variable function is defined by the n^2 partial derivatives:

$$\frac{\partial}{\partial x_i}\frac{\partial f(\mathbf{x})}{\partial x_j}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, n$$

written as:

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}, \quad i \neq j, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_i^2}, \quad i = j.$$

These n^2 second partial derivatives are usually represented by a square, symmetric matrix, termed the *Hessian matrix*, denoted by:

$$\mathbf{H}(\mathbf{x}) \equiv \nabla^2 f(\mathbf{x}) \equiv \begin{bmatrix} \dfrac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$

Example: For the previous example:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -x_2 \cos x_1 & 2x_2 - \sin x_1 \\ 2x_2 - \sin x_1 & 2x_1 \end{bmatrix}$$

Note: If the Hessian matrix of $f(\mathbf{x})$ is a constant matrix, $f(\mathbf{x})$ is then *quadratic*, expressed as:

$$f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \alpha$$

$$\Rightarrow \nabla f(\mathbf{x}) = \mathbf{H}\mathbf{x} + \mathbf{c}, \quad \Rightarrow \nabla^2 f(\mathbf{x}) = \mathbf{H}$$

Convex and Concave Functions

A function is called *concave* over a given region R if:

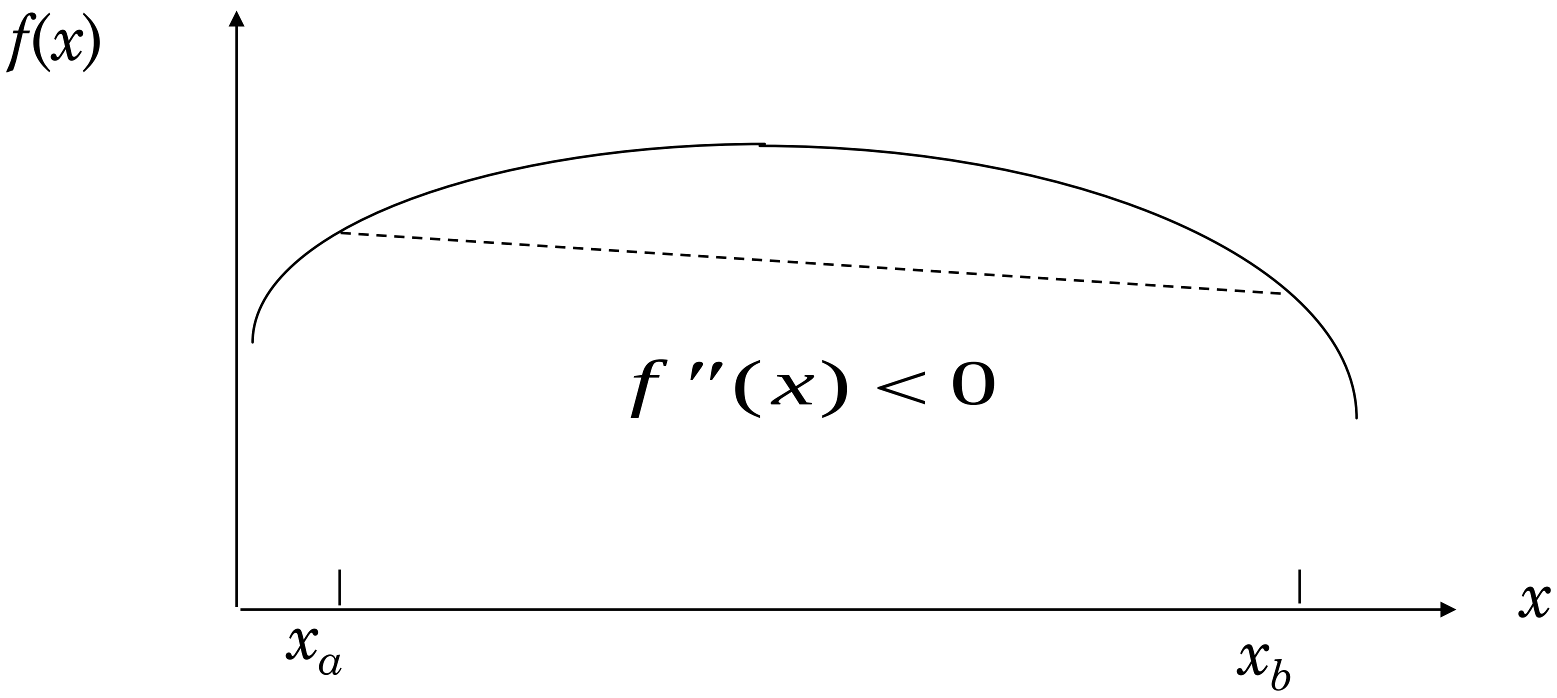
$$f(\theta \mathbf{x}_a + (1-\theta)\mathbf{x}_b) \geq \theta f(\mathbf{x}_a) + (1-\theta) f(\mathbf{x}_b)$$

$$\text{where: } \mathbf{x}_a, \mathbf{x}_b \in R, \quad \text{and } 0 \leq \theta \leq 1.$$

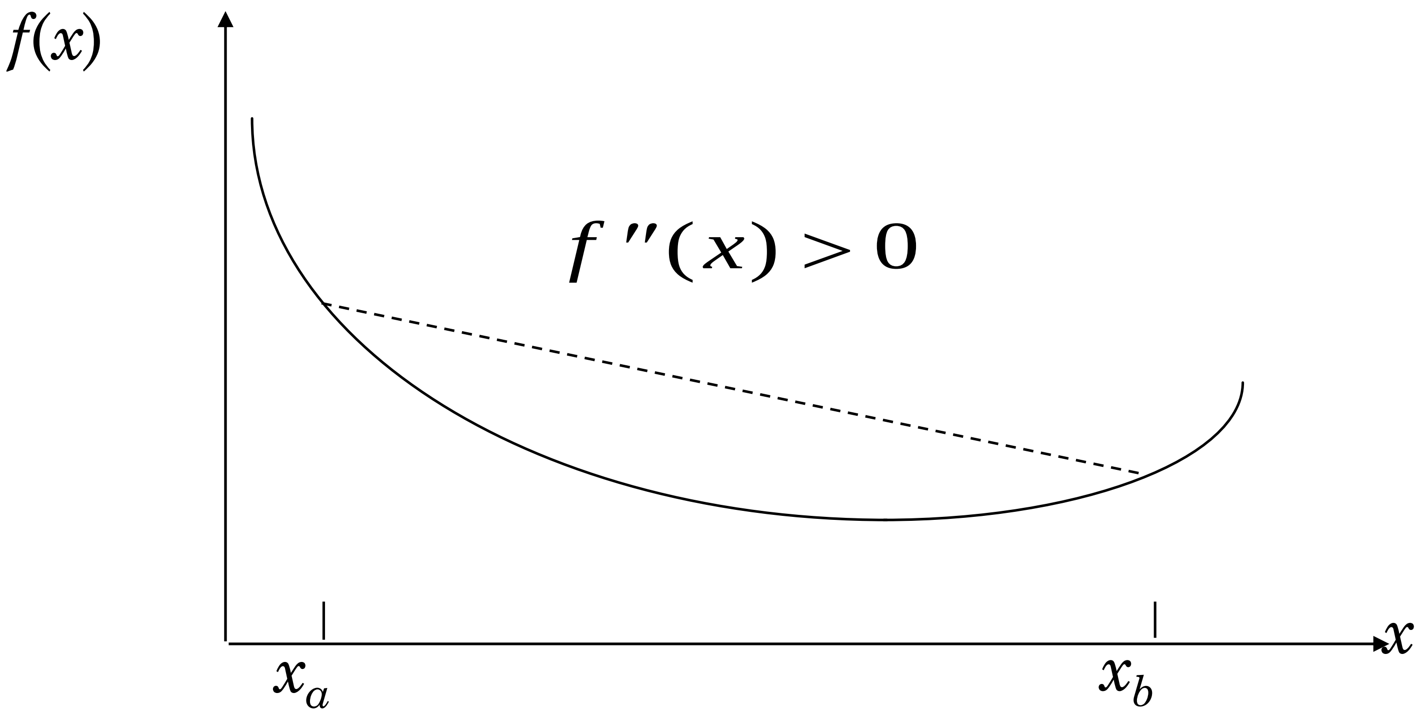
The function is *strictly concave* if $\geq$ is replaced by >.

A function is called convex (strictly convex) if $\geq$ is replaced by $\leq$ (<).

concave function



convex function



$$\text{If} \quad f''(x) = \frac{\partial^2 f}{\partial x^2} \leq 0 \quad \text{then } f(x) \text{ is concave.}$$

$$\text{If} \quad f''(x) = \frac{\partial^2 f}{\partial x^2} \geq 0 \quad \text{then } f(x) \text{ is convex.}$$

For a *multivariate function* $f(\mathbf{x})$ the conditions are:-

$f(\mathbf{x})$	$\mathbf{H}(\mathbf{x})$ *Hessian matrix*
Strictly convex	+ve def
convex	+ve semi def
concave	-ve semi def
strictly concave	-ve def

Tests for Convexity and Concavity

H is +ve def (+ve semi def) iff

$$\mathbf{x}^T\mathbf{H}\mathbf{x} > 0 \quad (\geq 0), \quad \forall \mathbf{x} \neq \mathbf{0}.$$

H is -ve def (-ve semi def) iff

$$\mathbf{x}^T\mathbf{H}\mathbf{x} < 0 \quad (\leq 0), \quad \forall \mathbf{x} \neq \mathbf{0}.$$

Convenient tests: **H**(**x**) is strictly convex (+ve def) (convex) (+ve semi def)) if:

1. all eigenvalues of **H**(**x**) are $> 0 \quad (\geq 0)$

or 2. all principal determinants of **H**(**x**) are $> 0 \quad (\geq 0)$

$\mathbf{H}(\mathbf{x})$ is strictly concave (-ve def) (concave (- ve semi def)) if:

1. all eigenvalues of $\mathbf{H}(\mathbf{x})$ are $< 0 \quad (\leq 0)$

or 2. the principal determinants of $\mathbf{H}(\mathbf{x})$ are alternating in sign:

$$\Delta_1 < 0,\ \Delta_2 > 0,\ \Delta_3 < 0,\ \cdots$$

$$(\Delta_1 \leq 0,\ \Delta_2 \geq 0,\ \Delta_3 \leq 0,\ \cdots)$$

Example

$$f(x) = 2x_1^2 - 3x_1x_2 + 2x_2^2$$

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 4x_1 - 3x_2 \qquad \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} = 4 \qquad \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} = -3$$

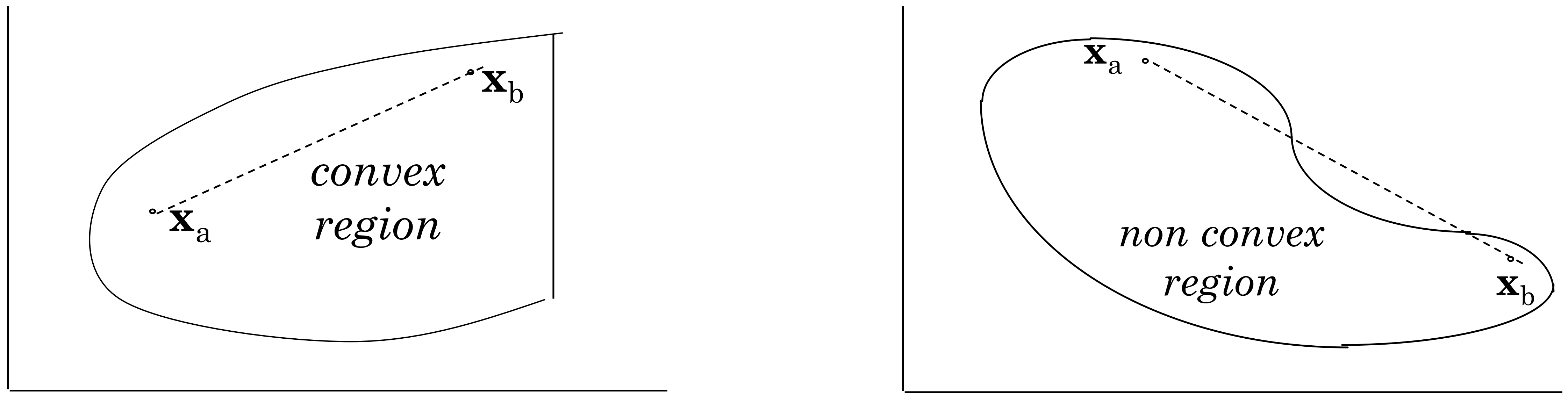
$$\frac{\partial f(\mathbf{x})}{\partial x_2} = -3x_1 + 4x_2 \qquad \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} = 4$$

$$\therefore \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}, \quad \Delta_1 = 4, \quad \Delta_2 = \begin{vmatrix} 4 & -3 \\ -3 & 4 \end{vmatrix} = 7$$

eigenvalues: $|\lambda \mathbf{I}_2 - \mathbf{H}| = \begin{vmatrix} \lambda - 4 & 3 \\ 3 & \lambda - 4 \end{vmatrix} = \lambda^2 - 8\lambda + 7 = 0$

$\Rightarrow \lambda_1 = 1,\ \lambda_2 = 7.$ Hence, $f(\mathbf{x})$ is strictly convex.

Convex Region



A convex set of points exist if for any two points, $\mathbf{x}_a$ and $\mathbf{x}_b$, in a region, all points:

$$\mathbf{x} = \mu \mathbf{x}_a + (1 - \mu)\mathbf{x}_b, \quad 0 \leq \mu \leq 1$$

on the straight line joining $\mathbf{x}_a$ and $\mathbf{x}_b$ are in the set.
If a region is completely bounded by concave functions then the functions form a convex region.

Necessary and Sufficient Conditions for an Extremum of an Unconstrained Function

A condition N is *necessary* for a result R if R can be true only if N is true.

$$R \Rightarrow N$$

A condition S is *sufficient* for a result R if R is true if S is true.

$$S \Rightarrow R$$

A condition T is *necessary and sufficient* for a result R iff T is true.

$$T \Leftrightarrow R$$

There are *two necessary and a single sufficient conditions* to guarantee that **x*** is an extremum of a function $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^*$:

1. $f(x)$ is twice continuously differentiable at x*.
2. $\nabla f(\mathbf{x}^*) = \mathbf{0}$, i.e. a stationary point exists at x*.
3. $\nabla^2 f(\mathbf{x}^*) = \mathbf{H}(\mathbf{x}^*)$ is +ve def for a minimum to exist at **x***, or -ve def for a maximum to exist at **x***

1 and 2 are necessary conditions; 3 is a sufficient condition.

Note: an extremum may exist at x* even though it is not possible to demonstrate the fact using the three conditions.

Example: Consider:

$$f(\mathbf{x}) = 4 + 4.5x_1 - 4x_2 + x_1^2 + 2x_2^2 - 2x_1x_2 + x_1^4 - 2x_1^2x_2$$

The gradient vector is:

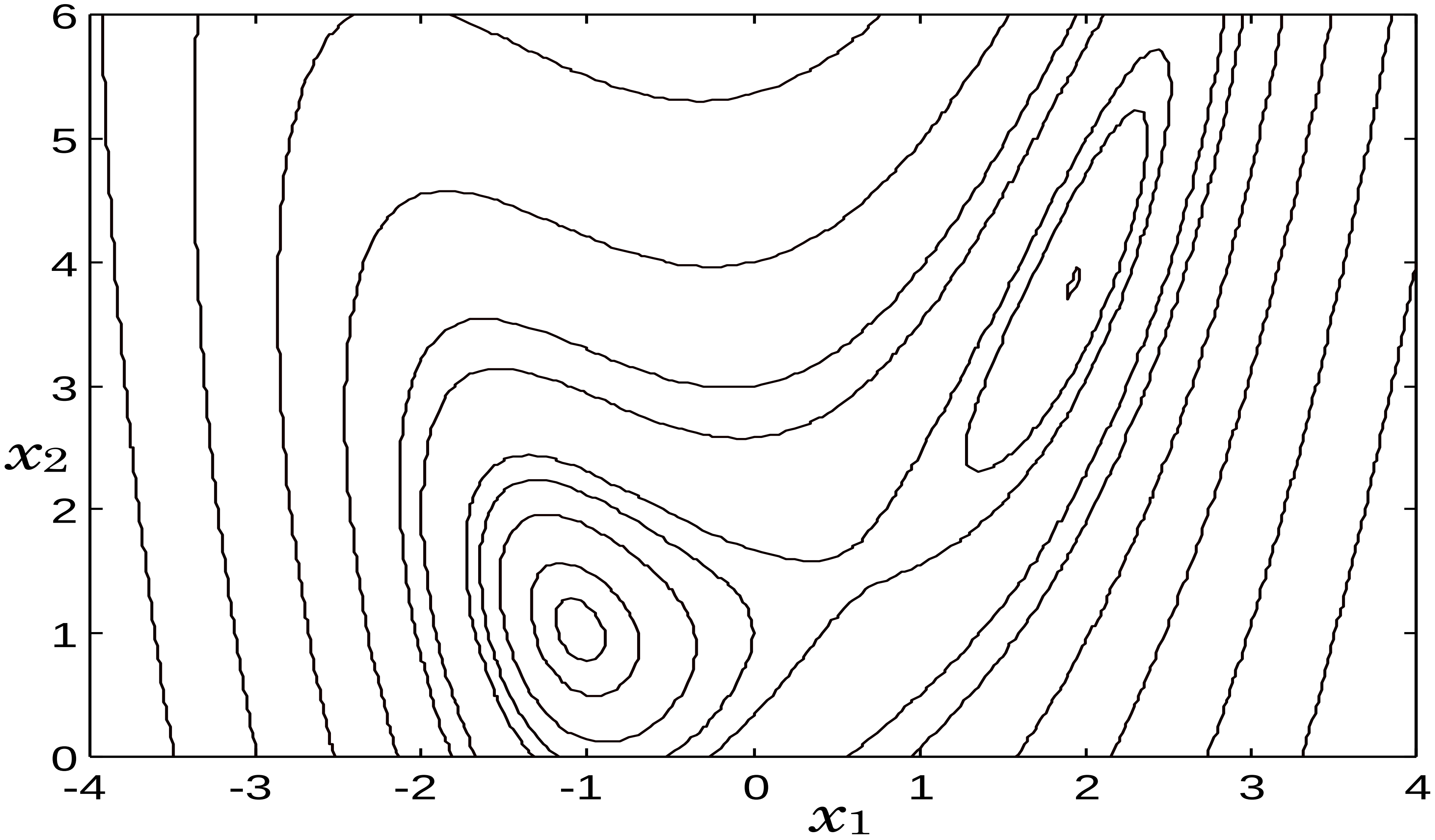
$$\nabla f(\mathbf{x}) = \begin{pmatrix} 4.5 + 2x_1 - 2x_2 + 4x_1^3 - 4x_1x_2 \\ -4 + 4x_2 - 2x_1 - 2x_1^2 \end{pmatrix}$$

yielding three stationary points located by setting $\nabla f(\mathbf{x}) = \mathbf{0}$ and solving numerically:

$\mathbf{x}^*=(x_1,x_2)$	$f(\mathbf{x}^*)$	eigenvalues of $\nabla^2 f(\mathbf{x})$	classification
A.(-1.05,1.03)	-0.51	10.5 3.5	*global min*
B.(1.94,3.85)	0.98	37.0 0.97	*local min*
C.(0.61,1.49)	2.83	7.0 -2.56	*saddle*

where: $$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 + 12x_1^2 - 4x_2 & -2 - 4x_1 \\ -2 - 4x_1 & 4 \end{pmatrix}$$

contour map



Interpretation of the Objective Function in Terms of its Quadratic Approximation

If a function of two variables can be approximated within a region of a stationary point by a *quadratic function*:

$$f(x_1, x_2) = \frac{1}{2}[x_1 \quad x_2]\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [c_1 \quad c_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \alpha$$

$$f(x_1, x_2) = \frac{1}{2}h_{11}x_1^2 + \frac{1}{2}h_{22}x_2^2 + h_{12}x_1x_2 + c_1x_1 + c_2x_2 + \alpha$$

then the *eigenvalues* and *eigenvectors* of:

$$\mathbf{H}(x_1^*, x_2^*) = \nabla^2 f(x_1^*, x_2^*) = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

can be used to interpret the nature of $f(x_1,x_2)$ at:

$$x_1 = x_1^*, \; x_2 = x_2^*$$

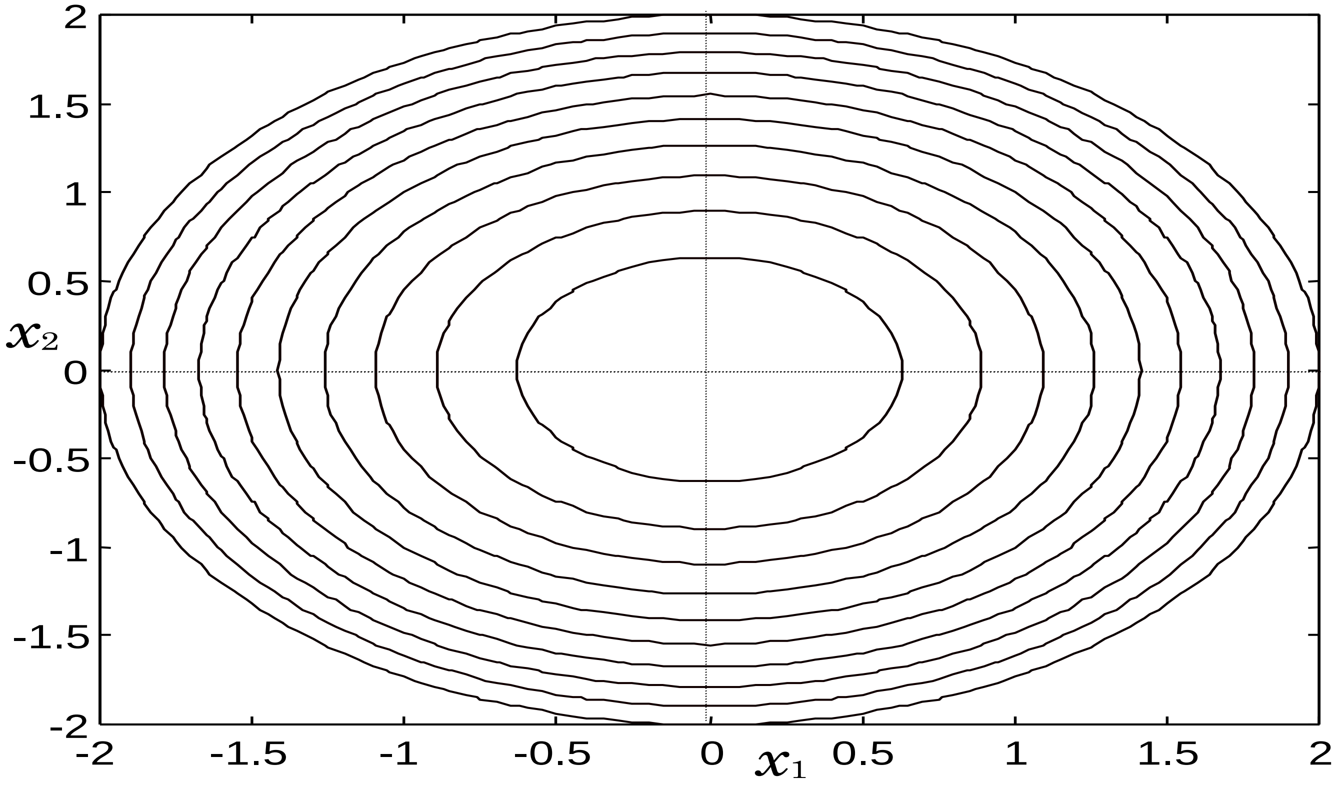
They provide information on the *shape* of $f(x_1,x_2)$ at $x_1 = x_1^*,\ x_2 = x_2^*$ If $\mathbf{H}(x_1^*, x_2^*)$ is +ve def, the eigenvectors are at right angles (*orthogonal*) and correspond to the principal axes of elliptical contours of $f(x_1,x_2)$.

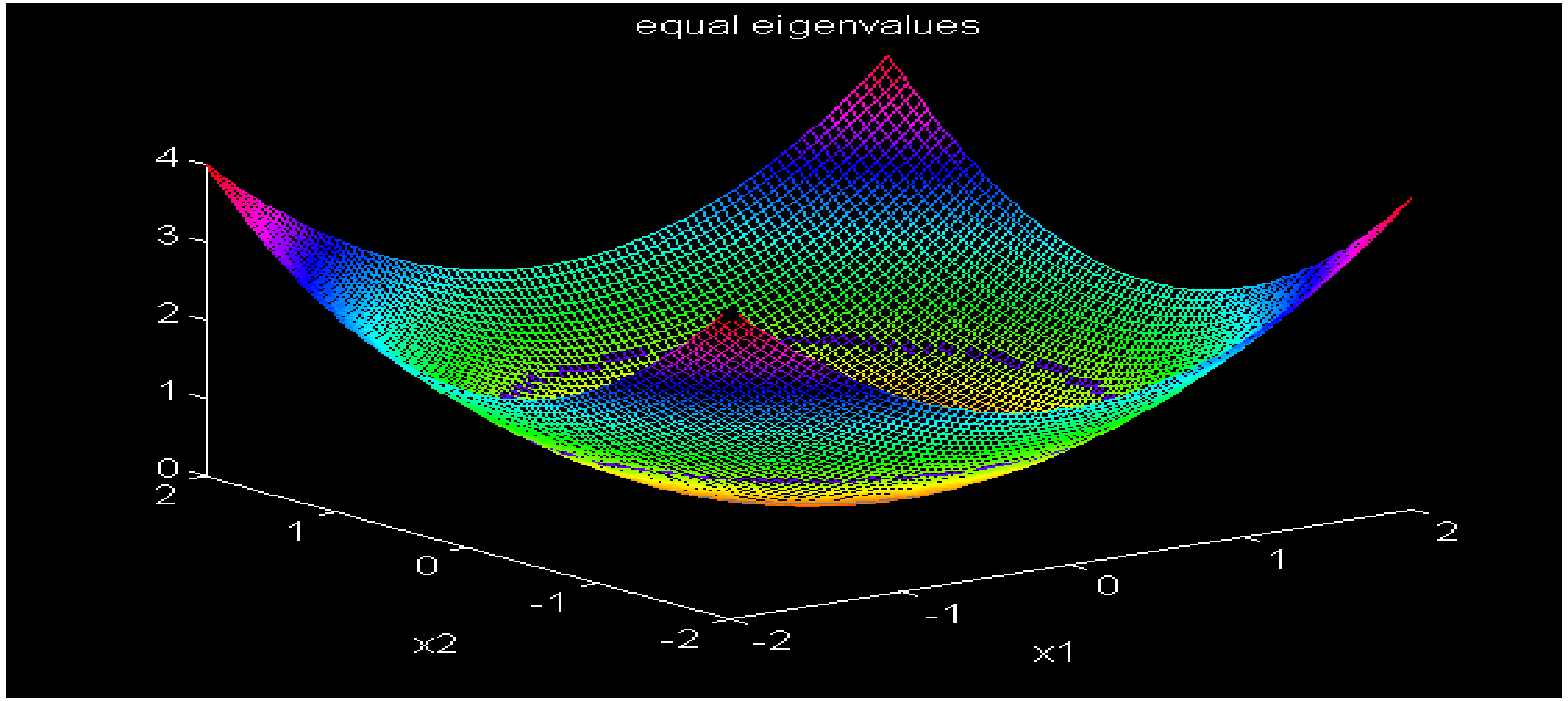
A valley or ridge lies in the direction of the eigenvector associated with a relative small eigenvalue.

These interpretations can be generalized to the *multivariate quadratic approximation*:

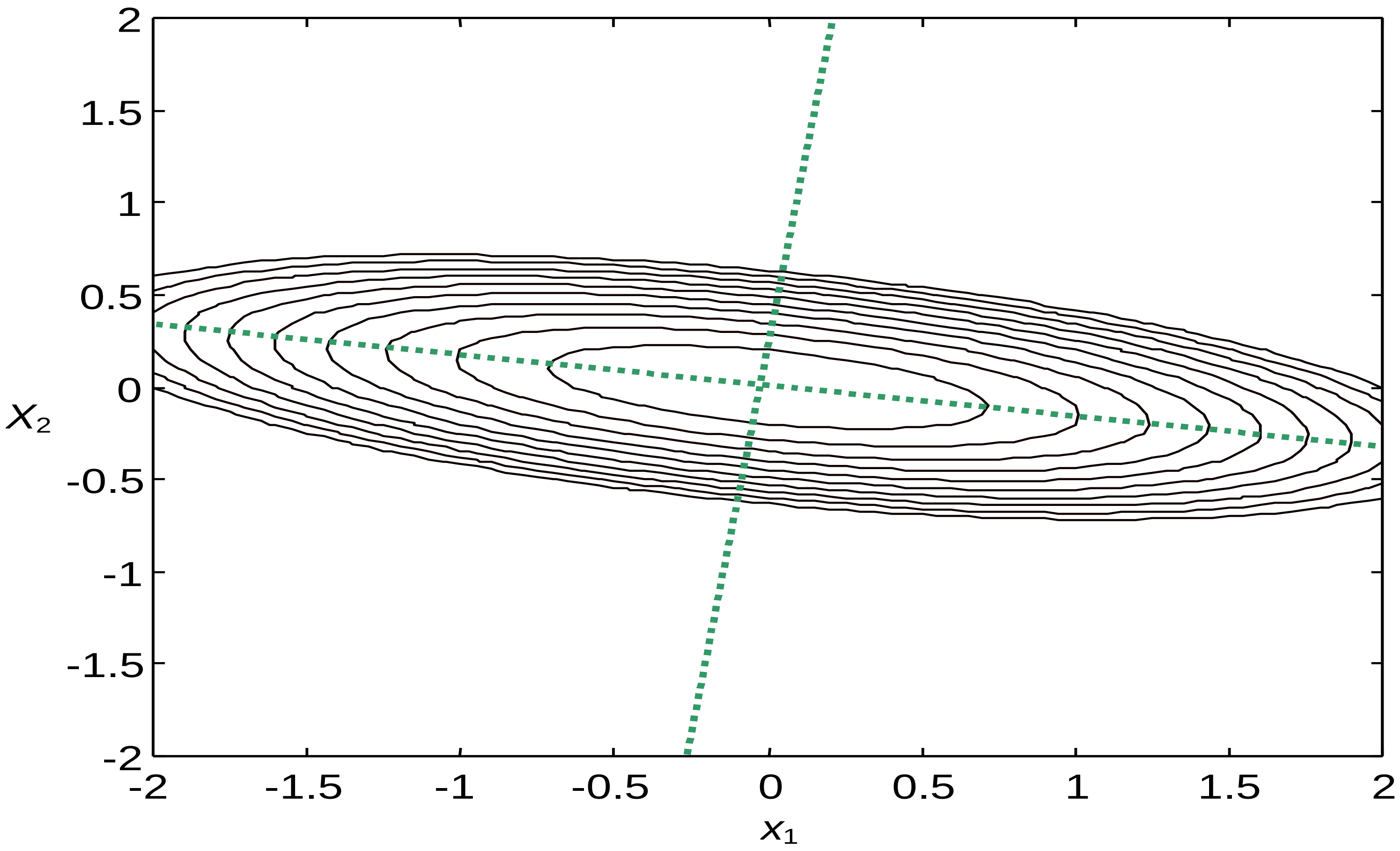
$$f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \alpha$$

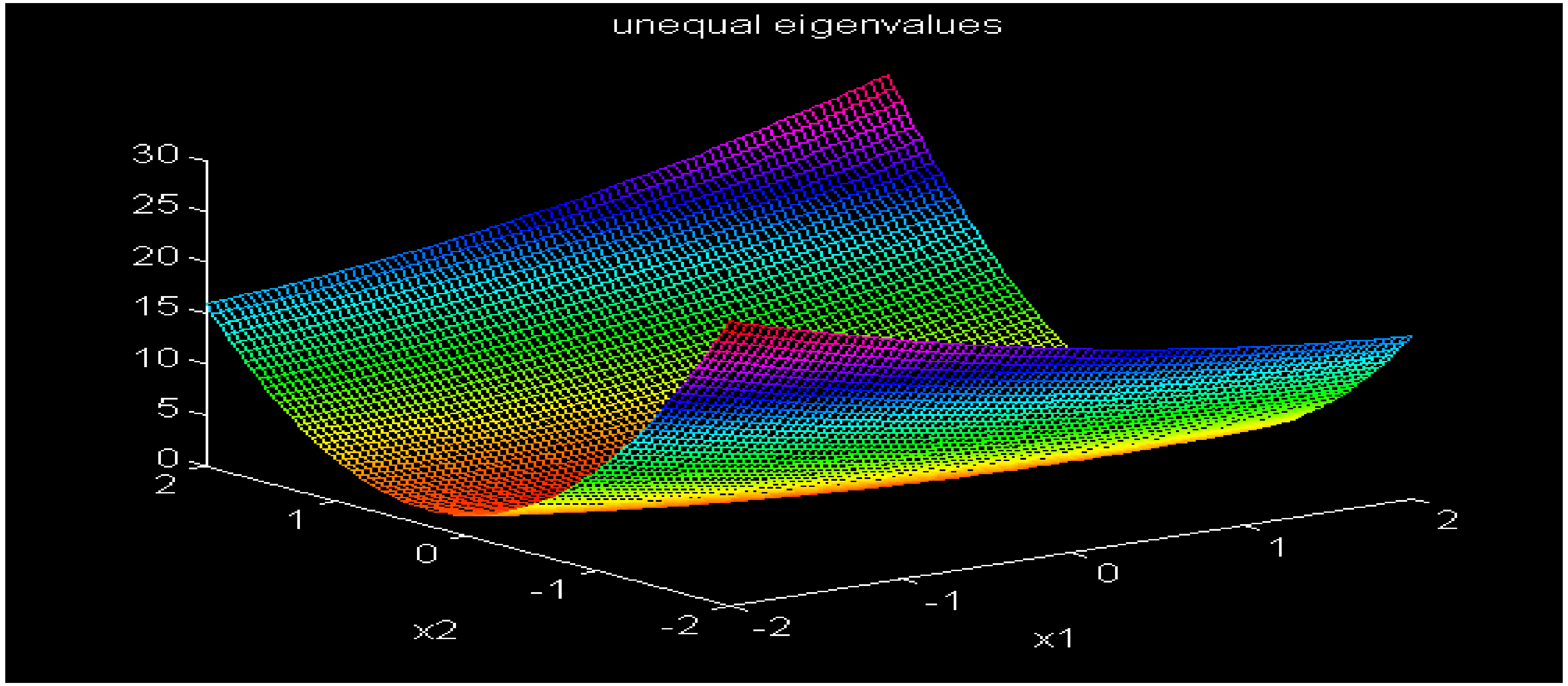
Case 1: *Equal Eigenvalues* - circular contours which are interpreted as a *circular hill* (max)(-ve eigenvalues) or circular valley (min) (+ve eigenvalues)



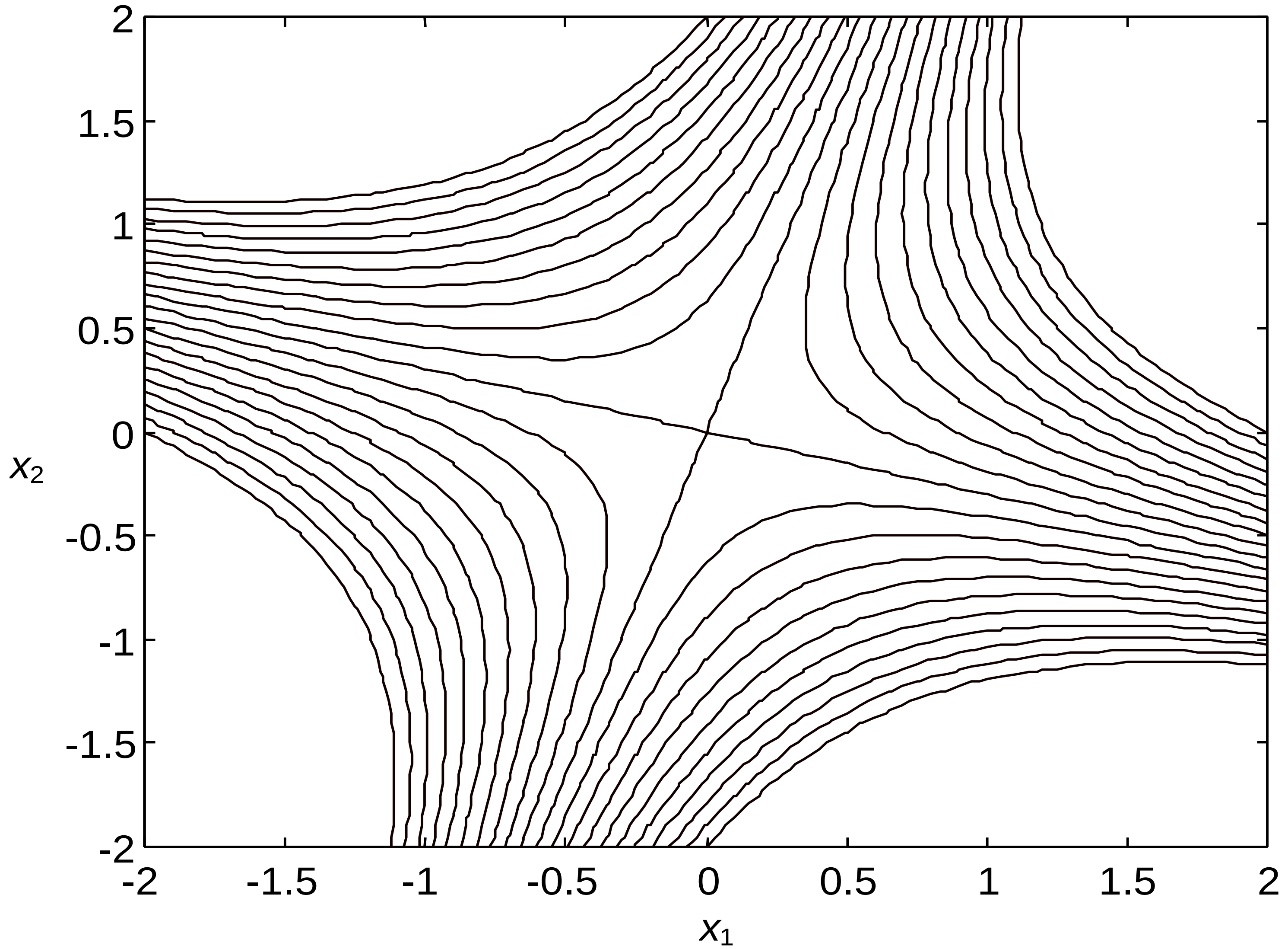

equal eigenvalues
4
3
2
1
0
2
1
0
-1
-2
x2
-2
-1
0
1
2
x1

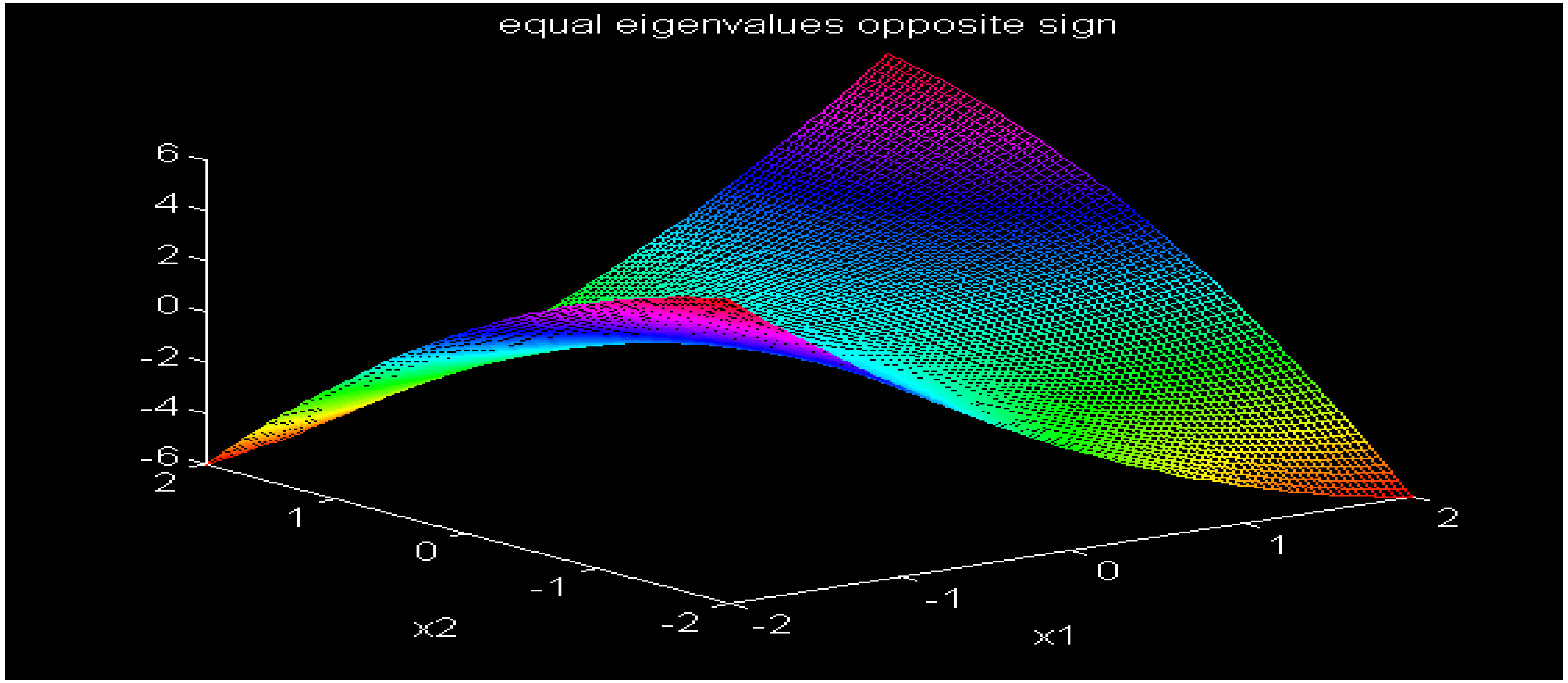
Case 2: *Unequal Eigenvalues (same sign)* – elliptical contours which are interpreted as an *elliptical hill* (max)(-ve eigenvalues) or elliptical valley (min)(+ve eigenvalues)



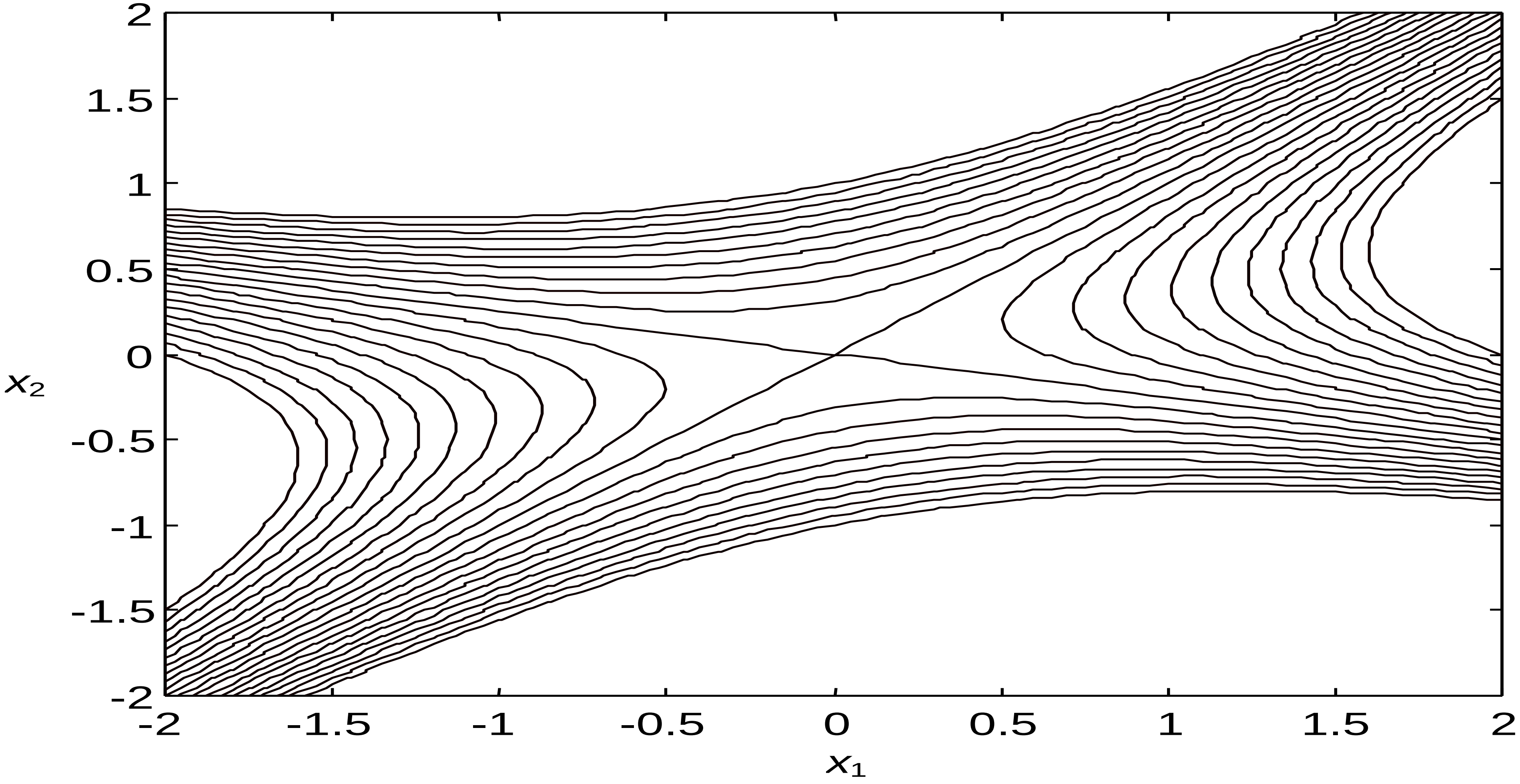

unequal eigenvalues
30
25
20
15
10
5
0
2
1
0
-1
-2
x2
-2
-1
0
1
2
x1

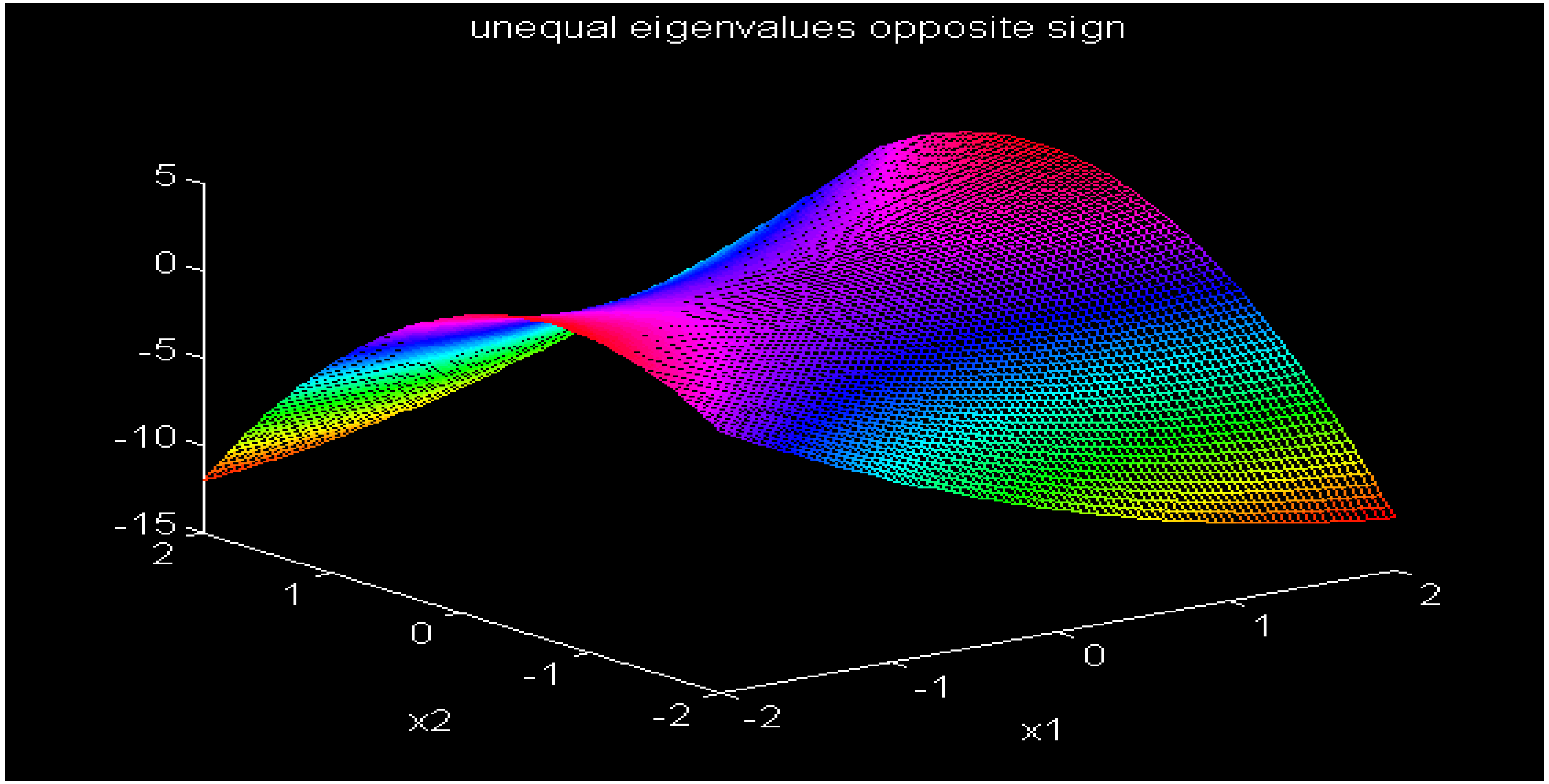
Case 3: *Eigenvalues of opposite sign but equal in magnitude* - *symmetrical saddle*.




equal eigenvalues opposite sign
6
4
2
0
-2
-4
-6
2
1
0
-1
-2
x2
-2
-1
0
1
2
x1

Case 4: *Eigenvalues of opposite sign but unequal in magnitude* - asymmetrical saddle.




unequal eigenvalues opposite sign
5
0
-5
-10
-15
2
1
0
-1
-2
x2
-2
-1
0
1
2
x1

Optimisation with Equality Constraints

$$\min_{\mathbf{x}} f(\mathbf{x}); \qquad \mathbf{x} \in \Re^{\mathbf{n}}$$

$$\text{subject to: } \mathbf{h}(\mathbf{x}) = \mathbf{0}; \quad \text{m constraints} \quad (\mathrm{m} \le \mathrm{n})$$

Elimination of variables:

example:

$$\min_{x_1, x_2} f(\mathbf{x}) = 4x_1^2 + 5x_2^2 \qquad \text{(a)}$$

$$\text{s.t. } 2x_1 + 3x_2 = 6 \qquad \text{(b)}$$

Using (b) to eliminate x_1 gives:

$$x_1 = \frac{6 - 3x_2}{2} \qquad \text{(c)}$$

and substituting into (a) :-

$$f(x_2) = (6 - 3x_2)^2 + 5x_2^2$$

At a stationary point

$$\frac{\partial f(x_2)}{\partial x_2} = 0 \rightarrow -6(6-3x_2)+10x_2 = 0$$

$$\rightarrow 28x_2 = 36 \rightarrow \underline{\underline{x_2^* = 1.286}}$$

Then using (c):

$$x_1^* = \frac{6-3x_2^*}{2} = \underline{\underline{1.071}}$$

Hence, the stationary point (min) is: (1.071, 1.286)

The Lagrange Multiplier Method

Consider a two variable problem with a single equality constraint:

$$\min_{x_1, x_2} f(x_1, x_2)$$
$$\text{s.t. } h(x_1, x_2) = 0$$

At a *stationary point* we may write:

$$\text{(a)} \quad df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0$$
$$\text{(b)} \quad dh = \frac{\partial h}{\partial x_1} dx_1 + \frac{\partial h}{\partial x_2} dx_2 = 0$$

If: $\begin{vmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \\ \dfrac{\partial h}{\partial x_1} & \dfrac{\partial h}{\partial x_2} \end{vmatrix} = 0$ nontrivial nonunique solutions for dx_1 and dx_2 will exist.

This is achieved by setting

$$\begin{vmatrix} -\lambda \dfrac{\partial h}{\partial x_1} & -\lambda \dfrac{\partial h}{\partial x_2} \\ \dfrac{\partial h}{\partial x_1} & \dfrac{\partial h}{\partial x_2} \end{vmatrix} \text{ with } \lambda = -\frac{\partial f}{\partial x_1} \Big/ \frac{\partial h}{\partial x_1} = -\frac{\partial f}{\partial x_2} \Big/ \frac{\partial h}{\partial x_2}$$

where λ is known as a *Lagrange multiplier*.

If an *augmented objective function*, called the ***Lagrangian*** is defined as:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2)$$

we can solve the constrained optimisation problem by solving:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda \frac{\partial h}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} + \lambda \frac{\partial h}{\partial x_2} = 0$$

provides equations (a) and (b)

$$\frac{\partial L}{\partial \lambda} = h(x_1, x_2) = 0$$

re-statement of equality constrain

Generalizing : To solve the problem:

$$\min_{\mathbf{x}} f(\mathbf{x}); \qquad \mathbf{x} \in \Re^{\mathbf{n}}$$

$$\text{subject to: } \mathbf{h}(\mathbf{x}) = \mathbf{0}; \quad m \text{ constraints } \ (m \le n)$$

define the Lagrangian:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}), \quad \lambda \in \Re^m$$

and the stationary point (points) is obtained from:-

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \left[\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \boldsymbol{\lambda} = \mathbf{0}$$

$$\nabla_{\lambda} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

Example Consider the previous example

again. The Lagrangian is:-

$$L = 4x_1^2 + 5x_2^2 + \lambda(2x_1 + 3x_2 - 6)$$

$$\frac{\partial L}{\partial x_1} = 8x_1 + 2\lambda = 0 \qquad \text{(a)}$$

$$\frac{\partial L}{\partial x_2} = 10x_2 + 3\lambda = 0 \qquad \text{(b)}$$

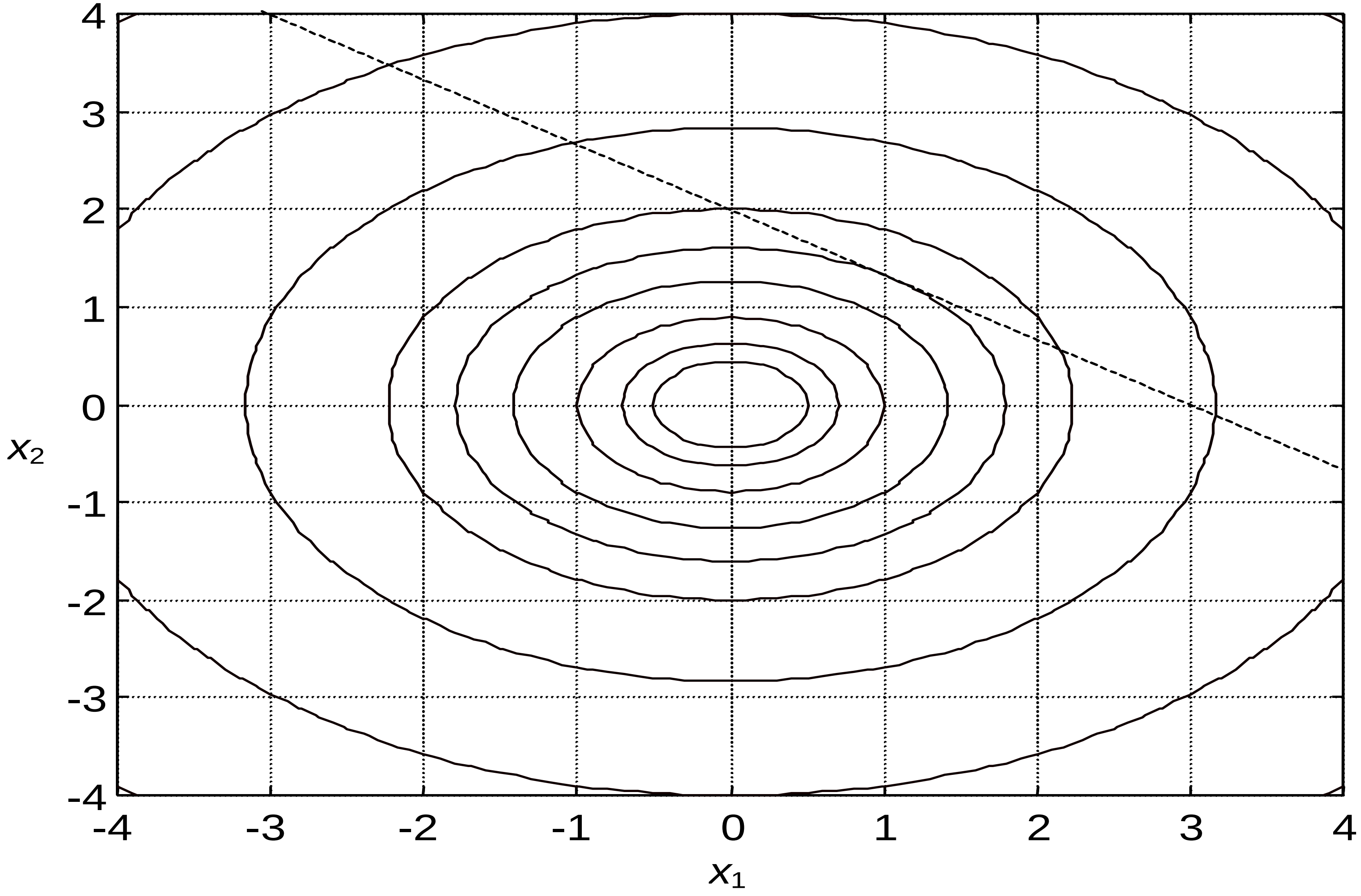
$$\frac{\partial L}{\partial \lambda} = 2x_1 + 3x_2 - 6 = 0 \qquad \text{(c)}$$

Substituting (a) and (b) into (c) gives:

$$x_1 = -\frac{\lambda}{4}, \quad x_2 = -\frac{3\lambda}{10} \quad \rightarrow -\frac{\lambda}{2} - \frac{9\lambda}{10} - 6 = 0 \rightarrow \lambda = -\frac{30}{7} = -4.281$$

$$\text{Hence,} \quad x_1 = \frac{15}{14} = \underline{\underline{1.071}}, \quad x_2 = \frac{90}{70} = \underline{\underline{1.286}}$$

which agrees with the previous result.



Necessary Conditions for a Local Extremum of an Optimisation Problem Subject to Equality and Inequality Constraints (Kuhn-Tucker Conditions)

Consider the problem:

$$\min_{\mathbf{x}} f(\mathbf{x}); \qquad \mathbf{x} \in \Re^{\mathbf{n}}$$

$$\text{s.t.:} \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}; \quad m \text{ equalities} \quad (m \le n)$$

$$\mathbf{g}(\mathbf{x}) \ge \mathbf{0}; \quad p \text{ inequalities}$$

Here, we define the *Lagrangian*:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) - \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}); \quad \boldsymbol{\lambda} \in \Re^m, \quad \boldsymbol{\mu} \in \Re^p$$

The *necessary conditions* for $\mathbf{x}^*$ to be a local extremum of $f(\mathbf{x})$ are:-

(a) $f(\mathbf{x})$, $h_j(\mathbf{x})$, $g_j(\mathbf{x})$ are all twice differentiable at $\mathbf{x}^*$

(b) The Lagrange multipliers exist

(c) All constraints are satisfied at $\mathbf{x}^*$

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0} \rightarrow h_j(\mathbf{x}^*) = 0; \quad \mathbf{g}(\mathbf{x}^*) \geq \mathbf{0} \rightarrow g_j(\mathbf{x}^*) \geq 0$$

(d) The Lagrange multipliers μ_j^* (at $\mathbf{x}^*$) for the inequality constraints are not negative, i.e.

$$\mu_j^* \geq 0$$

(e) The binding (active) inequality constraints are zero, the inactive inequality constraints are > 0, and the associated μ_j's are 0 at $\mathbf{x}^*$, i.e.

$$\mu_j^* g_j(\mathbf{x}^*) = 0$$

(f) The Lagrangian function is at a stationary point $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0$

Notes:

1. Further analysis or investigation is required to determine if the extremum is a minimum (or maximum)

2. If $f(\mathbf{x})$ is convex, $\mathbf{h}(\mathbf{x})$ are linear and $\mathbf{g}(\mathbf{x})$ are concave:

   $\mathbf{x}^*$ will be a *global extremum*.

Limitations of Analytical Methods

The computations needed to evaluate the above conditions can be extensive and intractable. Furthermore, the resulting simultaneous equations required for solving $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ are often nonlinear and cannot be solved without resorting to numerical methods.

The results may be inconclusive.

For these reasons, we often have to resort to numerical methods for solving optimisation problems, using computer codes (e.g.. MATLAB)

Example

Determine if the potential minimum
x* = (1.00,4.90) satisfies the Kuhn Tucker conditions for the problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) = 4x_1 - x_2^2 - 12$$

$$\text{s.t. } h_1(\mathbf{x}) = 25 - x_1^2 - x_2^2 = 0$$

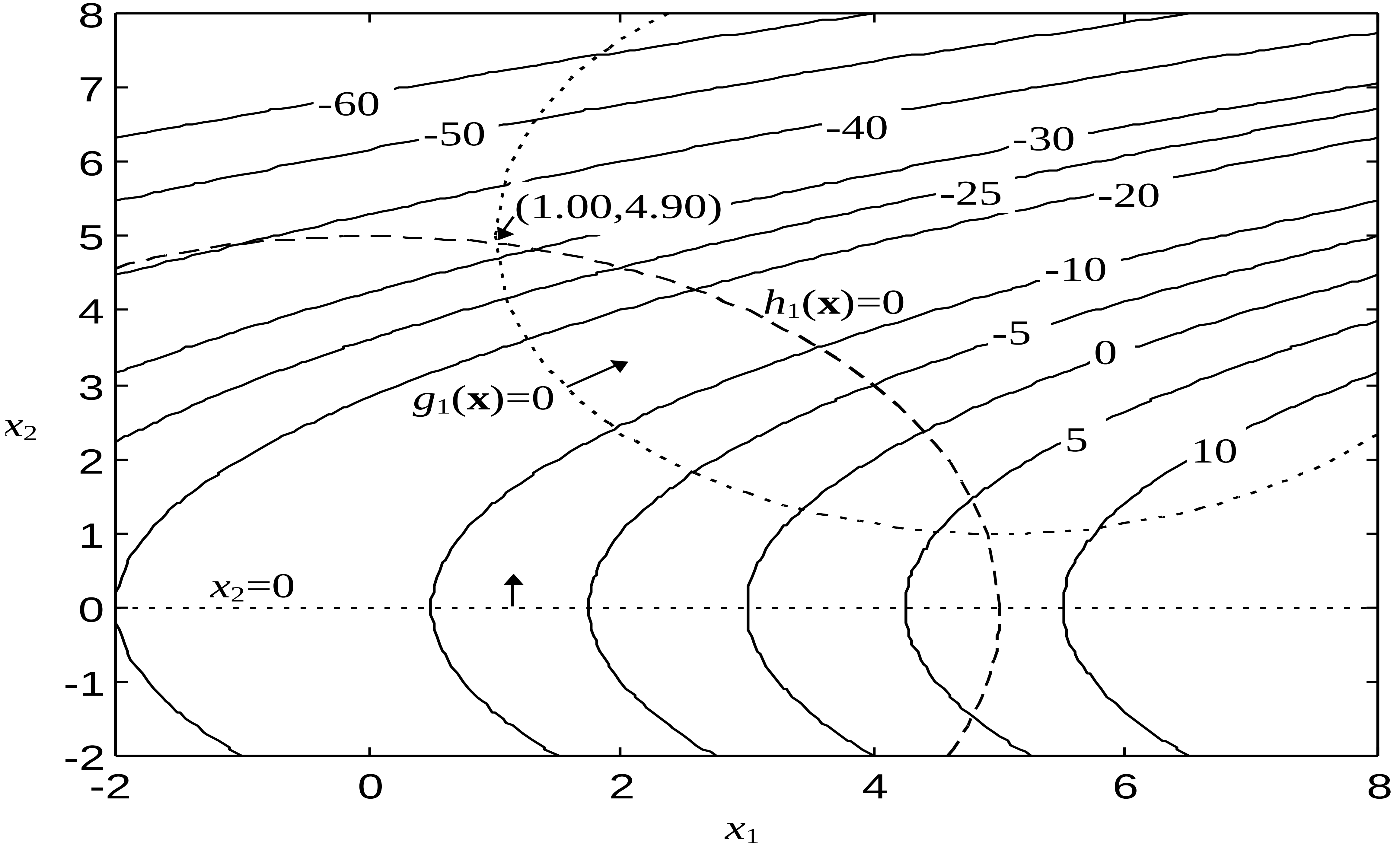
$$g_1(\mathbf{x}) = 10x_1 - x_1^2 + 10x_2 - x_2^2 - 34 \geq 0$$

$$g_2(\mathbf{x}) = (x_1 - 3)^2 + (x_2 - 1)^2 \geq 0$$

$$g_3(\mathbf{x}) = x_1 \geq -2$$

$$g_4(\mathbf{x}) = x_2 \geq 0$$

contours and constraints



We test each Kuhn-Tucker condition in turn:

(a) All functions are seen by inspection to be twice differentiable.

(b) We assume the Lagrange multipliers exist

(c) Are the constraints satisfied?

$$h_1: \quad 25-(1.00)^2-(4.90)^2=-0.01\approx 0 \quad \textit{yes}$$

$$g_1: \quad 10(1.00)-(1.00)^2+10(4.90)-(4.90)^2-34=-0.01\approx 0 \quad \textit{yes binding}$$

$$g_2: \quad (1.00-3)^2+(4.90-1)^2=19.21\geq 0 \quad \textit{yes not active}$$

$$g_3: \quad 1.00\geq -2 \quad \textit{yes not active}$$

$$g_4: \quad 4.90\geq 0 \quad \textit{yes not active}$$

To test the rest of the conditions we need to determine the Lagrange multipliers using the stationarity conditions. First we note that from condition (e) we require:

$$\mu^*_j g_j(\mathbf{x}^*) = 0, \quad j = 1,2,3,4$$

μ_1^* can have any value because $g_1(\mathbf{x}^*) = 0$

μ_2^* must be zero because $g_2(\mathbf{x}^*) > 0$

μ_3^* must be zero because $g_3(\mathbf{x}^*) > 0$

μ_4^* must be zero because $g_4(\mathbf{x}^*) > 0$

Now consider the stationarity condition:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \quad \text{where, since } \mu_2^* = \mu_3^* = 0$$

$$L = 4x_1 - x_2^2 - 12 + \lambda_1(25 - x_1^2 - x_2^2) - \mu_1(10x_1 - x_1^2 + 10x_2 - x_2^2 - 34)$$

Hence:

$$\nabla_{x_1} L = 4 - 2\lambda_1^* x_1^* - \mu_1^*(10 - 2x_1^*) = 4 - 2\lambda_1^* - 8\mu_1^*$$

$$\nabla_{x_2} L = -2x_2^* - 2\lambda_1^* x_2^* - \mu_1^*(10 - 2x_2^*) = -9.8 - 9.8\lambda_1^* - 0.2\mu_1^*$$

$$\begin{bmatrix} 2 & 8 \\ 9.8 & 0.2 \end{bmatrix} \begin{bmatrix} \lambda_1^* \\ x_1^* \end{bmatrix} = \begin{bmatrix} 4 \\ -9.8 \end{bmatrix} \qquad \lambda_1^* = -1.015, \ \mu_1^* = 0.754$$

Now we can check the remaining conditions:

(d) Are $\mu_j^* \geq 0,\ j = 1,2,3,4$?

$$\mu_1^* = 0.754,\ \mu_2^* = \mu_3^* = \mu_4^* = 0$$

Hence, the answer is *yes*

(e) Are $\mu_j^* g_j(\mathbf{x}^*) = 0,\ \ j = 1,2,3,4$?

Yes, because we have already used this above.

(f) Is the Lagrangian function at a stationary point?

Yes, because we have already used this above.

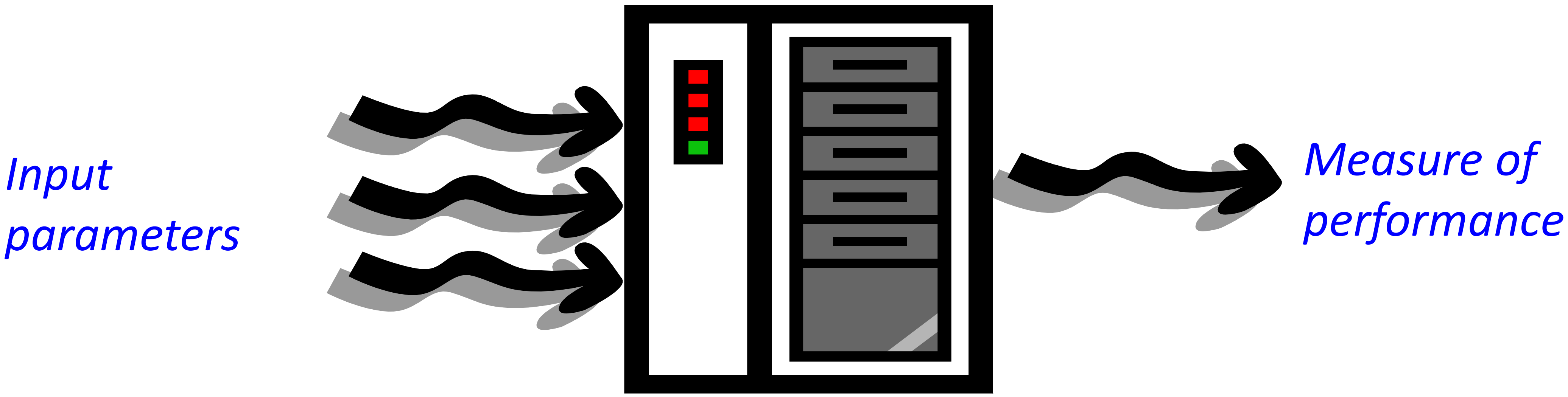
Hence, all the Kuhn-Tucker conditions are satisfied and we can have confidence in the solution :

$$x^* = (1.00, 4.90)$$

Simulation Optimization

What is Simulation Optimization?

- The optimization of simulation models deals with the situation in which the analyst would like to find which of possibly many sets of model specifications (i.e., input parameters and/or structural assumptions) leads to optimal performance



Simulation Optimization
Why is it required?

- Complex models contain many variables and constraints as well as uncertainty
- What-if simulation analysis unlikely to result in an optimal answer due to large number of possible solutions
- Inability of pure optimization to model complexities, uncertainties and dynamics of scenarios.
- Simulation-Optimization removes these inabilities by combining both approaches.

Simulation-Optimization
Why is it required?

- A total solution requires both capabilities.

- Two-Step Solution
  - Simulation
  - Optimization

- Both are necessary, neither is sufficient.

Simulation Optimization
Benefits in Dealing with Uncertainty

- Simulation enables understanding/modeling and communications of uncertainty.

- Optimization enables management of uncertainty.

Classical Approaches

- Stochastic approximation
  - Gradient-based approaches
- Sequential response surface methodology
- Random search
- Sample path optimization
  - Also known as stochastic counterpart