به نام خدا

BSS

Hw-1

محمدرضا آرانی

810100511

دانشگاه تهران

1401/12/08

جدول محتویات

بخش-1:

سوال-1:

بخش اول

منبع s_1 را که یک فرآیند تصادفی دارای توزیع یکنواخت بین [3 -3] می باشد با T=1000 نمونه و منبع s_2 را که یک فرآیند تصادفی دارای توزیع یکنواخت بین [2 -2] می باشد با T=1000 نمونه تولید کنید. این دو منبع را به صورت خطی و آنی توسط ماتریس مخلوط کننده $A = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}$ ترکیب کنید و مشاهدات x_1 و x_2 را تولید کنید.

۱) نمودار پراکندگی x_2 را بر حسب x_1 رسم کنید (scatter plot). رابطه ی این شکل با ماتریس A چیست؟

در این سوال، پس از تولید نمونه‌ها، آنها را با ماتریس A مخلوط کرده و به ماتریس مشاهدات می‌رسیم.

در عمل داریم:

$$x_1(t) = A_{11}s_1(t) + A_{12}s_2(t);$$
$$x_2(t) = A_{21}s_1(t) + A_{22}s_2(t);$$

با در نظر گرفتن بردارهای مکانی a_2 و a_1 به صورت زیر:

$$a_1 = [A_{11}; A_{21}];$$
$$a_2 = [A_{12}; A_{22}];$$

مشاهدات ما به فرم زیر در خواهد آمد:

$$X_{2*T} = A_{2*2}S_{2*T} = a_1s_1(t) + a_2s_2(t)\,;$$

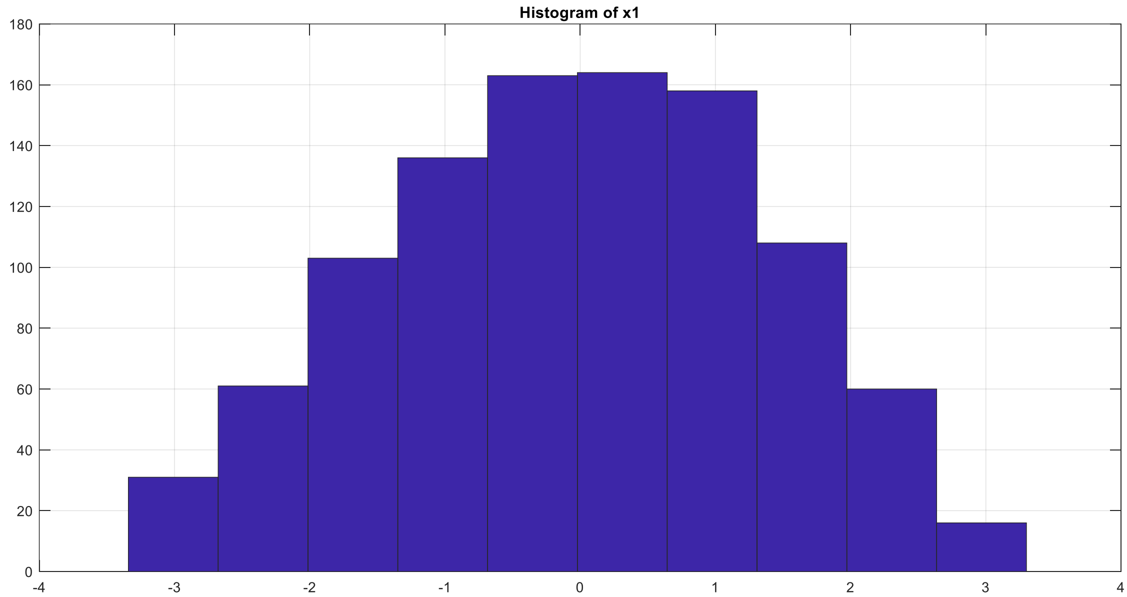
در واقع هردوی x_2 و x_1 حاصل ترکیب خطی از نمونه‌های یکنواخت اند که منجر به توزیعی شبیه به توزیع Irwin-Hall که درواقع جمع تعدادی فرآیند یکنواخت مستقل است، می‌شود.

در اینجا ضرایب وزن متفاوتی به هر منبع توسط ماتریس A که همان کانال ما بوده، نسبت داده می‌شود.

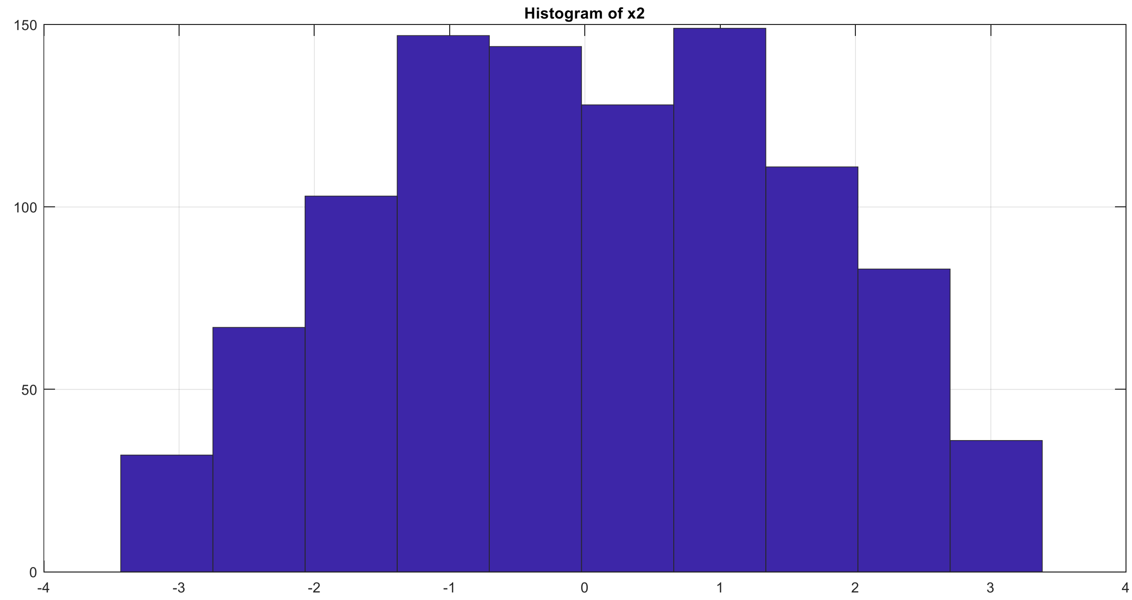
توزیع جدید باید از روی کانولوشن توزیع‌های وزن‌دار s_1 و s_2 به دست آید.

با توجه به اینکه توزیع s_1 و s_2 یکنواخت می‌باشد، توزیع هر یک از مشاهدات نیز باید تقریبا به صورت مثلثی که همان کانولوشن دو پالس مربعی می‌باشد درآید.

در شکل زیر توزیع‌های x1 و x2 را خواهیم دید:

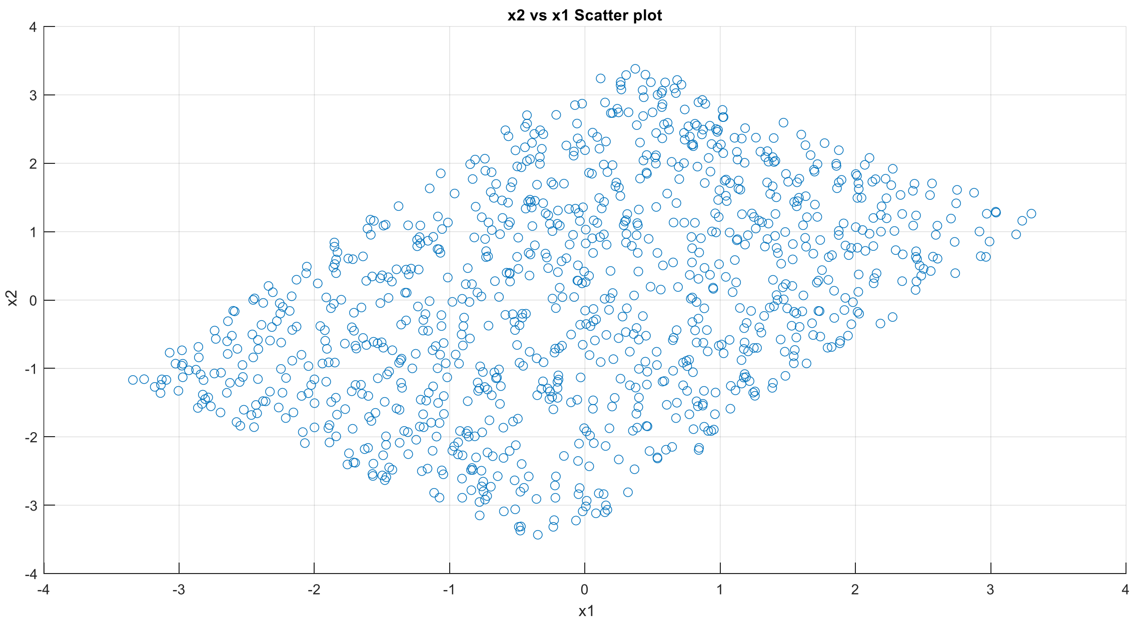


شکل 1



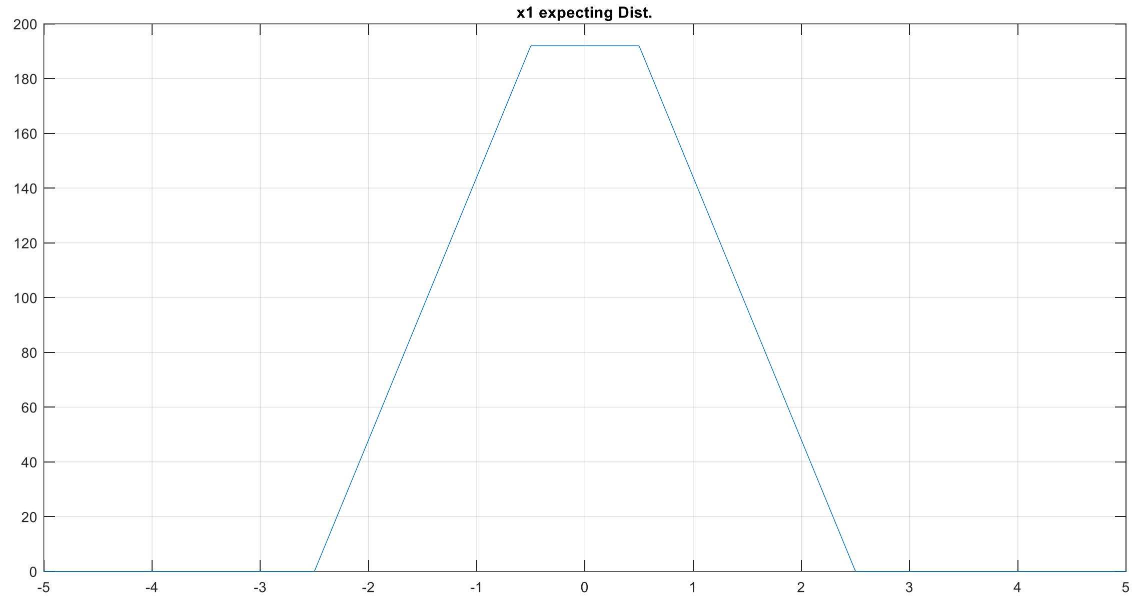
شکل 2

توزیع داده‌ها به صورت زیر خواهد بود:

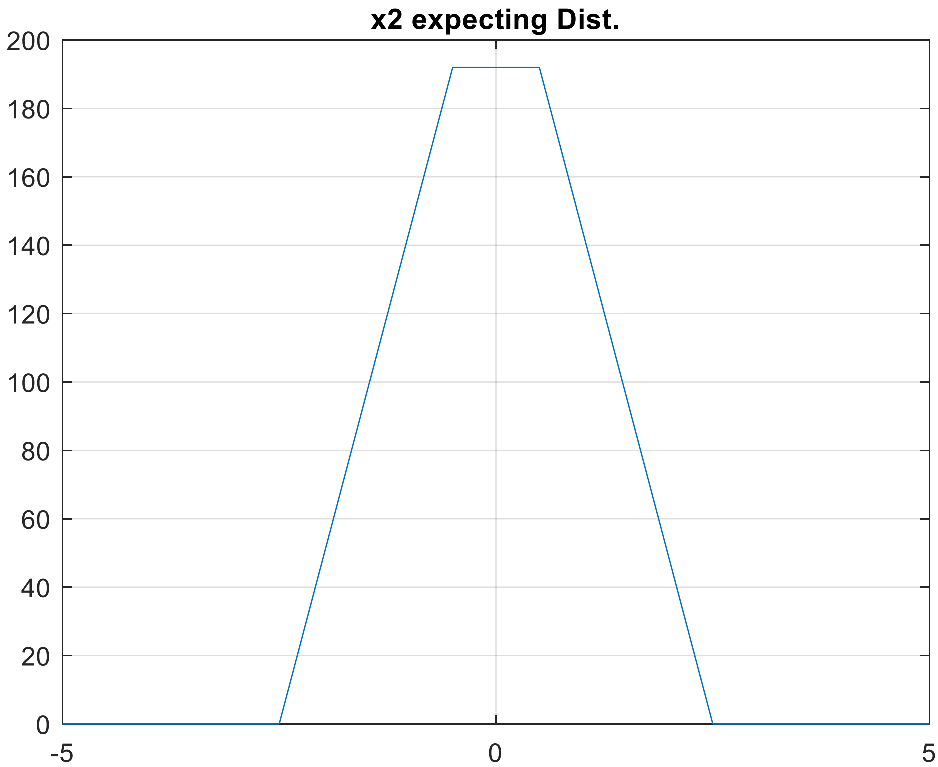


شکل 3

توزیع داده‌های مورد انتظار به صورت زیر است:



شکل 4



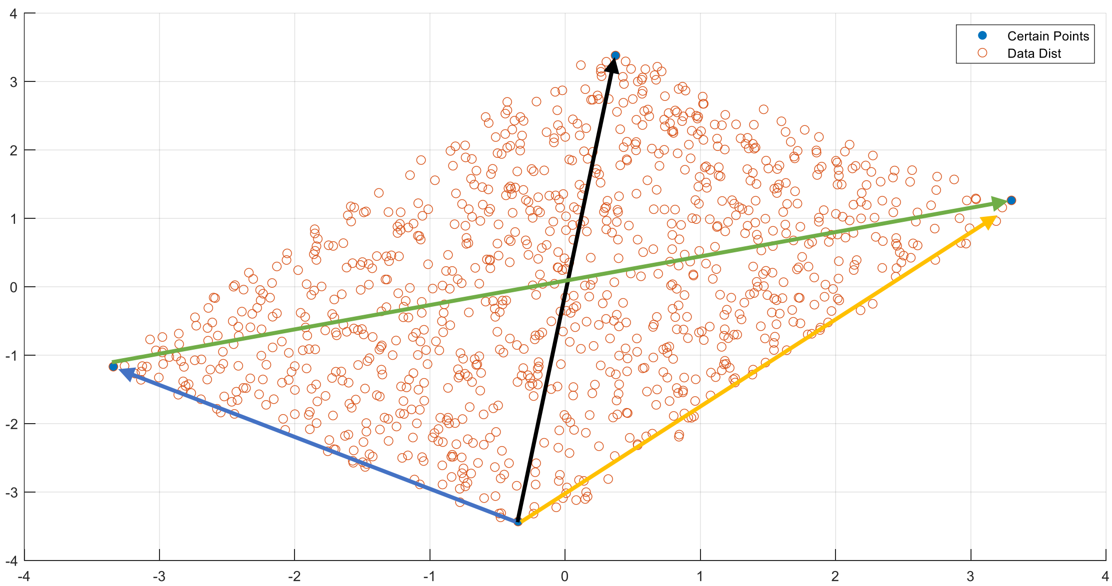
شکل 5

سوال-2:

۲) یک روش ریاضی (مرتبط با پاسخ قسمت ۱) پیشنهاد دهید که به صورت اتوماتیک ماتریس A را در این مساله بیابیم. روش خود را پیاده سازی کنید.

برای پیداکردن ماتریس A نیاز هست که بردارهای a_2 و a_1 را به درستی انتخاب کنیم. برای پیداکردن درست این بردارها، بهترین کار استفاده از PCA بوده که بهترین محور را برای توزیع موجود جهت داشتن بیشترین تفکیک پذیری انتخاب می‌کند. حال از آنجا که فرض بر آن است که نه درس ML پاس کردیم و نه پردازش آرایه‌ای، می‌توان حاشیه‌های این داده‌ها پیداکرده و آنها را به عنوان بردارهای متناسب انتخاب کنیم.

در اینجا می‌توان از شکل هندسی حاصله استفاده کرد و از روی جمع و تفریق دو محور مورد نظر، نقاطمورد علاقه را پیداکرد و سپس از روی سیستم دو معادله دو مجهول، a_1, a_2 را به دست آورد.



شکل 6

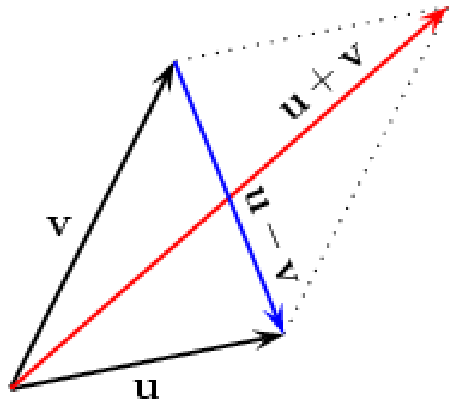
در شکل فوق، محور زرد، a_1 و محور آبی a_2 می‌باشد.

روش پیشنهادی برای پیداکردن اتوماتیک a_2 و a_1:

در شکل فوق، خط مشکی، نشان دهنده‌ی $a_1 + a_2$ و خط سبز نشان‌دهنده‌ی $a_1 - a_2$ می‌باشد.

$$Diff = a_1 - a_2;$$

$$sum = a_1 + a_2;$$



7 شکل

$$sum + Diff = 2 * a_1;$$

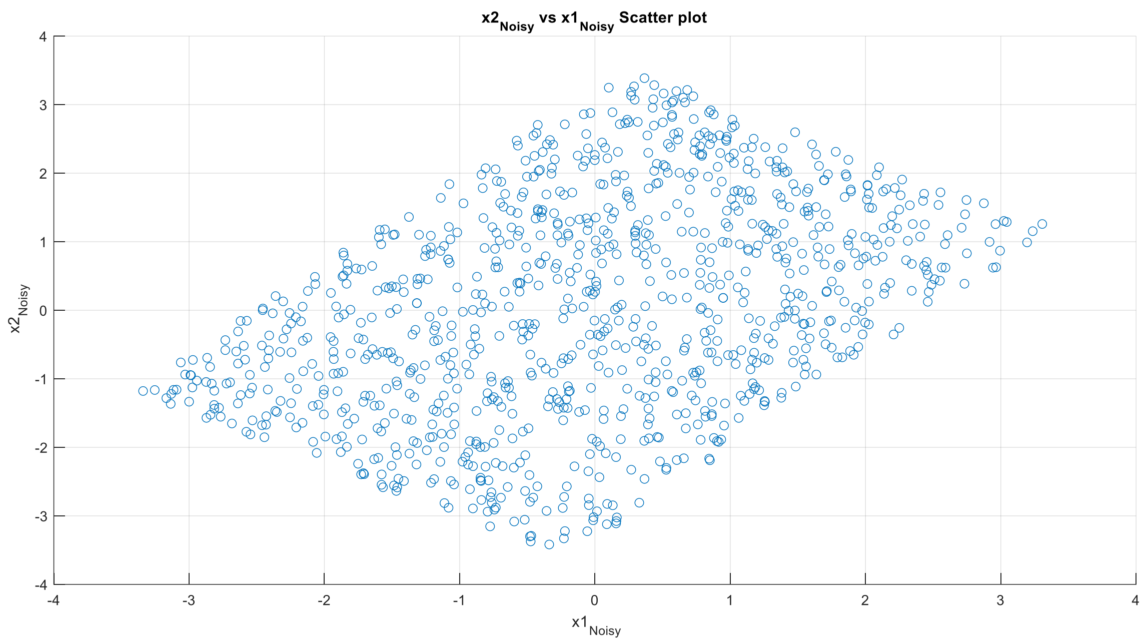
$$sum - Diff = 2 * a_2;$$

سوال-3:

۳) حال به ماتریس مشاهدات نویز (با قدرت پایین) اضافه کنید و روش پیشنهادی در قسمت قبل را روی آن اعمال کنید. آیا روش شما در سناریوی جدید نیز قابل پیاده سازی است؟ اگر خیر، روشتان را تغییر دهید. عملکرد روشتان را در حضور نویز گزارش دهید. متری پیشنهاد دهید که نشان دهد روش شما در حضور نویز نیز عملکرد مناسبی دارد.

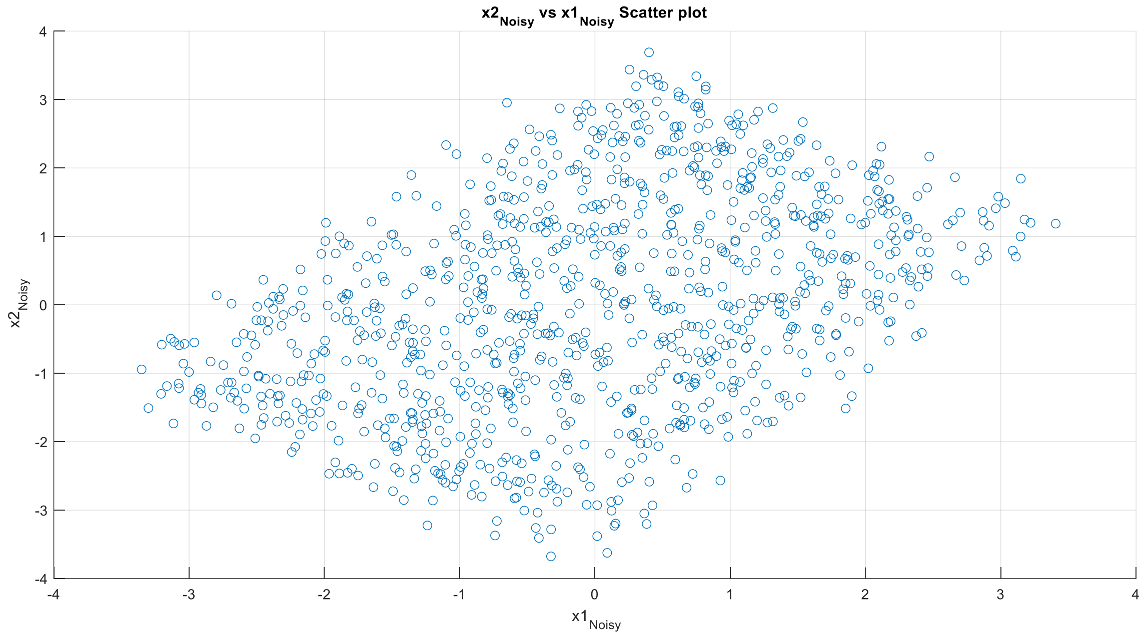
ابتدا نقاط ابتدا و انتها را مانند شکل قبل به دست آورده، سپس یک خط تشکیل می‌شود و ما نقاط نزدیک به این خط را شناسایی کرده و سپس آنها را در یک ست جمع آوری کرده و بر این ست جدید یک خط برازش خواهیم کرد. خط fit شده باید نسبت به داده‌های نویزی نیز خوب عمل کند.

در حضور نویز داده‌ها شکل زیر را به خود خواهند گرفت:



شکل 8

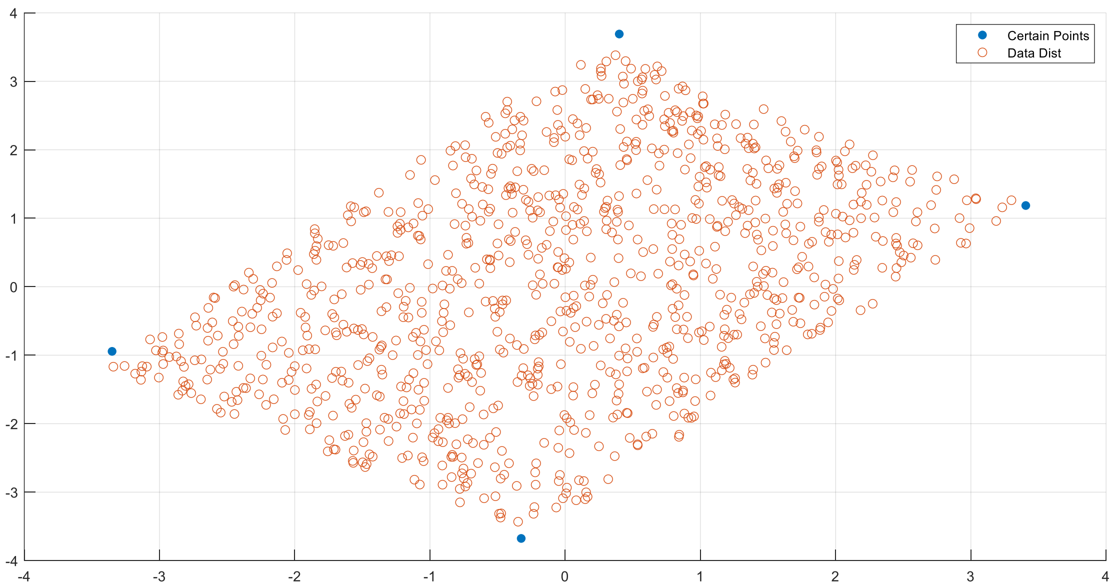
نویز گاوسی با واریانس 0.1



شکل 9

نویز گاوسی با واریانس 0.5

حال نقاط را مانند قبل انتخاب می‌کنیم:



شکل 10

مشاهده می‌شود که نقاط کمی با جای ایده‌آل خود فاصله دارند:

روش همچنان به درستی عمل می‌کند و مقادیر a_2 , a_1 در هر دو حالت تقریبا یکسان اند.

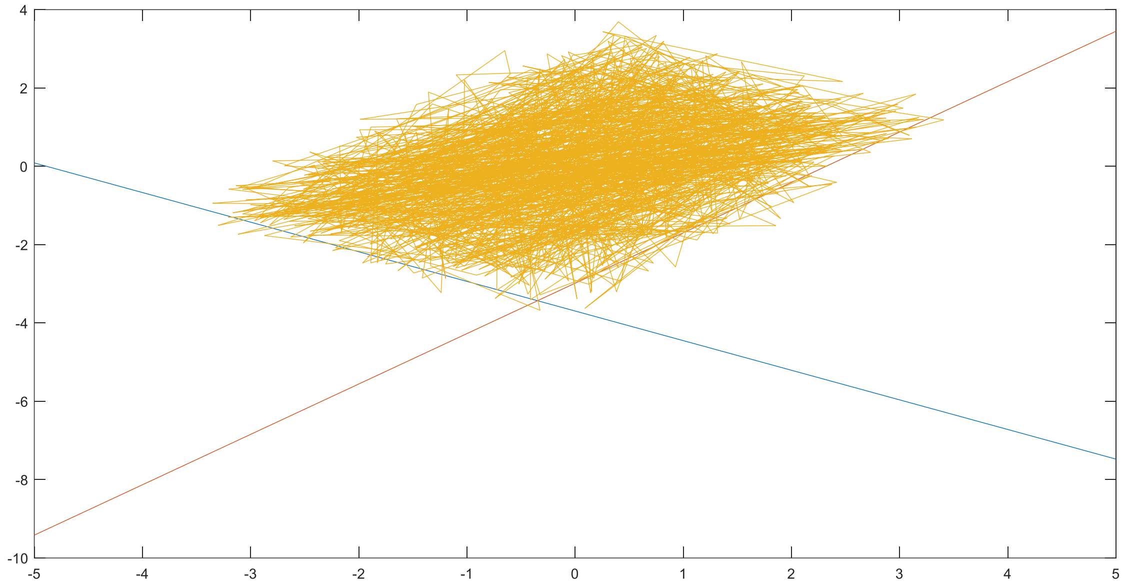
```
>> a_2_orginal_Noisy

a_2_orginal_Noisy =

   -2.9930    2.2637

>> a_2_orginal

a_2_orginal =

   -2.9930    2.2637
```

شکل 11

```
>> a_1_orginal_Noisy

a_1_orginal_Noisy =

    3.6483    4.6947

>> a_1_orginal

a_1_orginal =

    3.6483    4.6947
```

شکل 12

پس از پیدا کردن خط، به خروجی زیر خواهیم رسید:



شکل 13

که تقریبا به درستی این خطوط شناسایی شده اند.

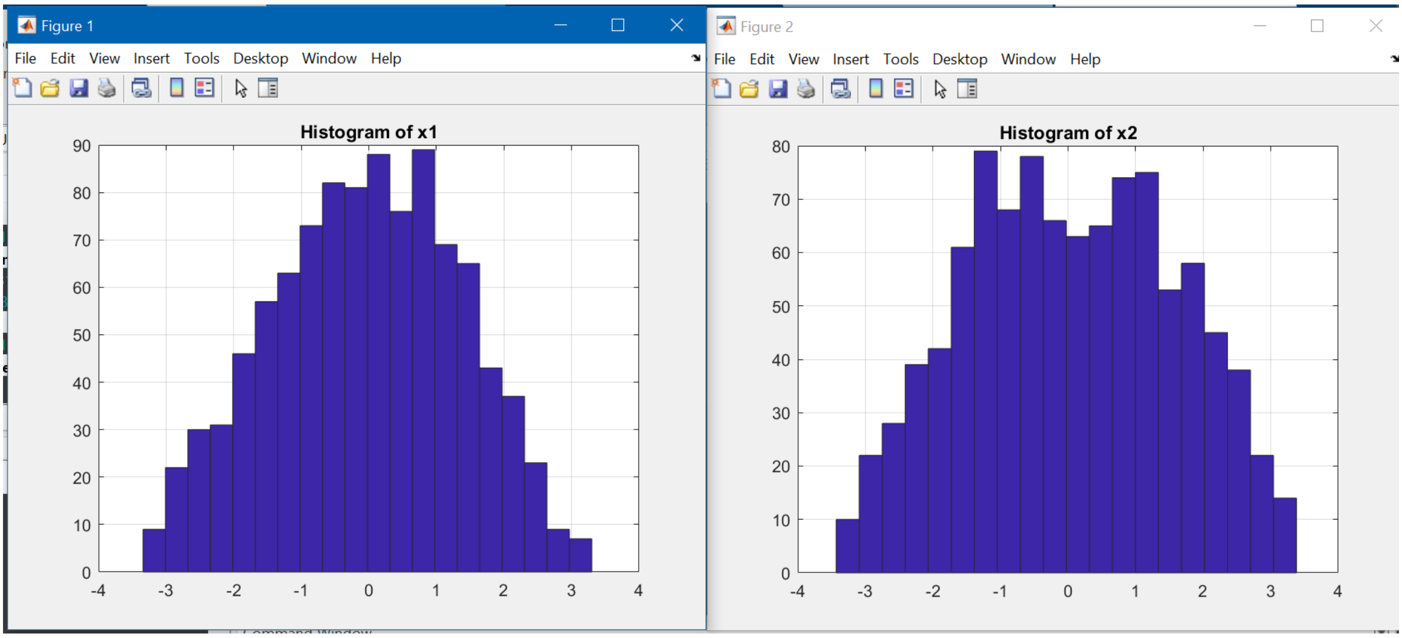
برای پیشنهاد دادن متری که نشان دهد روش ما در حضور نویز به درستی کار می‌کند، باید به این مسئله به صورت یک مسئله‌ی classification نگاه کرد که در آن اگر داده ها در سمت درستی از خط قرار گرفته بودن، آنها را درست لیبل بزند و اگر خیر لیبل غاط بزند و در نهایت تعداد درست را به تعداد کل تقسیم نموده و درصد دقت را به دست آورد.

سوال-4,5.

۴) هیستوگرام x_1 را با در نظر گرفتن 20 بین (bin) رسم کنید و با استفاده از روابط آماری و ریاضی تابع توزیع x_1 را هم به دست آورید. تابع توزیعی که به دست آوردید باید با هیستوگرام تطابق داشته باشد.

۵) قسمت ۴ را برای x_2 تکرار کنید.

در قسمت‌های قبل توضیح داده شد.



شکل 14

در واقع هردوی x_2 و x_1 حاصل ترکیب خطی از نمونه‌های یکنواخت اند که منجر به توزیعی شبیه به توزیع Irwin-Hall که درواقع جمع تعدادی فرآیند یکنواخت مستقل است، می‌شود.

در اینجا ضرایب وزن متفاوتی به هر منبع توسط ماتریس A که همان کانال ما بوده، نسبت داده می‌شود.

توزیع جدید باید از روی کانولوشن توزیع‌های وزن‌دار s_1 و s_2 به دست آید.

با توجه به اینکه توزیع s_1 و s_2 یکنواخت می‌باشد، توزیع هر یک از مشاهدات نیز باید تقریبا به صورت مثلثی که همان کانولوشن دو پالس مربعی می‌باشد درآید.

```
%%

Ts=1;
fs=1/Ts;
fsample=100;

Tsample=1/fsample;
t=-5: Tsample :5; % time vector

TAU1 = 6 ;
TAU2 = 4 ;

h1=heaviside(t+TAU1/2)-heaviside(t-TAU1/2);
h2=heaviside(t+TAU2/2)-heaviside(t-TAU2/2);

New_Dist=conv(0.8*h1,-0.6*h2);
figure()
plot(linspace(-5,5,length(New_Dist)),abs(New_Dist));
grid on
title("x2 expecting Dist.")
```

شکل 15

سوال-6.

۶) فرض کنید T به اندازه کافی بزرگ باشد و بعد از کشیدن هیستوگرام ها (قسمت های ۴ و ۵) شکل ها دقیقا مطابق تابع توزیع هایی باشند که با استفاده از روابط آماری و ریاضی به دست می آید. همچنین فرض کنید توزیع دقیق منابع (یکنواخت بودنشان + بازه ی مقادیرشان) را هم می دانیم. آیا در این حالت با استفاده از هیستوگرام ها می توان مساله BSS را حل کرد؟ توضیح دهید.

با داشتن دانش نسبت به بازه و مقادیر از روی هیستوگرام، می‌توان به صورت زیر داشت:

$$s_1 \sim U(p,q)$$

$$s_2 \sim U(m,n)$$

$$x_1 = A_{11}s_1 + A_{12}s_2 =>$$

$$x_1 \sim U(A_{11}p, A_{11}q) * U(A_{12}m, A_{12}n);$$

با داشتن 4 معادله و 2 مجهول، می‌توان با داشتن بازه‌ی داده‌ها، از روی هیستوگرام مقادیر A_{12} و A_{11} را پیداکرد.

با ادامه‌ی همین فرآیند برای x_2 می‌توان مقادیر A_{21}, A_{22} را پیدا کرد.

بخش-2:

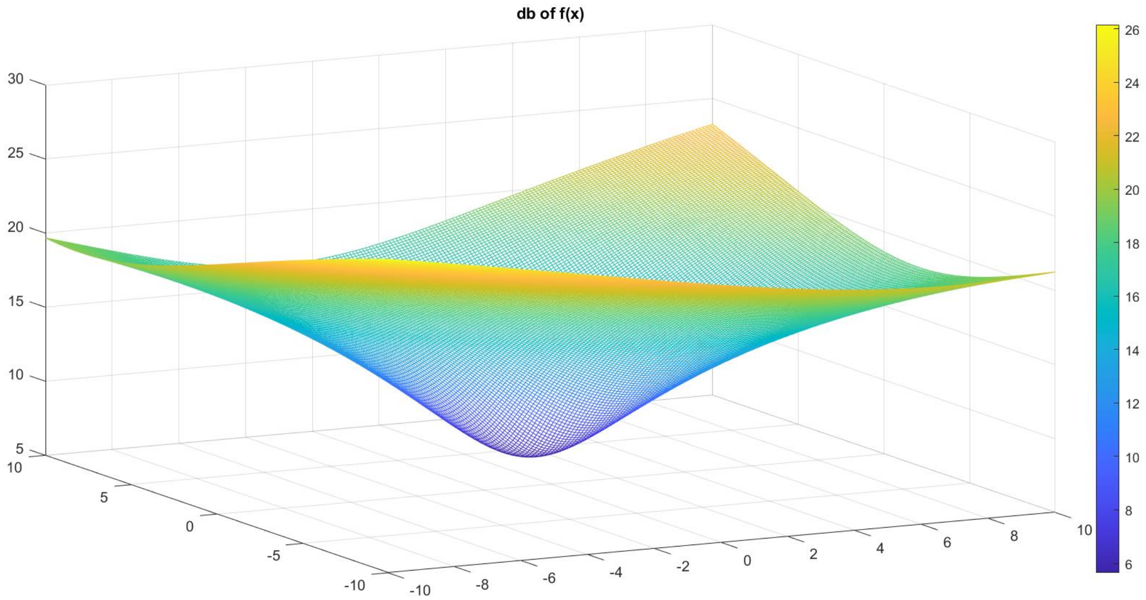
سوال-1:

تابع هدف دو متغیره زیر را در نظر بگیرید

$$f(\underset{\sim}{x}) = x_1^2 + x_2^2 - 4x_1 - 6x_2 + 13 + x_1x_2$$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

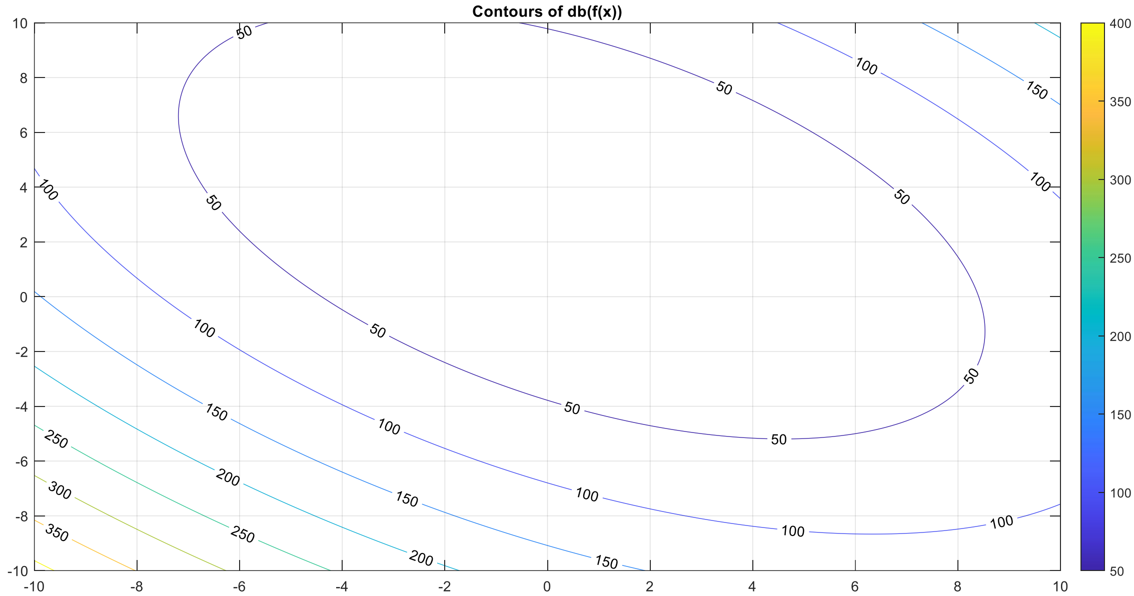
۱) با استفاده از دستور mesh این تابع را بر حسب x_1 و x_2 در بازه ی $-10 < x_2, x_1 < 10$ به صورت سه بعدی رسم کنید. بهتر است تابع هدف را به صورت db رسم کنید تا تغییرات تابع هدف برایتان واضح تر شود.



شکل 16

سوال-2:

۲) با استفاده از دستور contour سطوح هم پتانسیل این تابع را در فضای دو بعدی رسم کنید. مقادیر سطوح هم پتانسیل نیز روی شکل نشان داده شود (help دستور را مطالعه کنید).



شکل 17

سوال-3:

۳) بردار گرادیان (g) و ماتریس هسین (H) را محاسبه کنید. با توجه به ماتریس هسین به دست آمده، ثابت کنید تابع هدف داده شده convex است و در نتیجه فقط یک مینیمم دارد. (همان طور که در کلاس اشاره شد باید ثابت کنید نامساوی $\underline{x}^T \ H \ \underline{x} > 0$ به ازای همه ی مقادیر برداری <u>x</u> برقرار است.)

برای محاسبه‌ی بردار گرادیان و هسین داریم:

$$gf = \begin{bmatrix} \frac{\partial}{\partial} x_1 \\ \frac{\partial}{\partial} x_2 \end{bmatrix} f = \begin{bmatrix} 2x_1 - 4 + x_2 \\ 2x_2 - 6 + x_1 \end{bmatrix}$$

$$Hf = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_2 x_1} \\ \frac{\partial^2}{\partial x_2 x_1} & \frac{\partial^2}{\partial x_2^2} \end{bmatrix} f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

برای اثبات اینکه این تابع convex است، باید ماتریس هسین آن، PD و یا PSD باشد.

اگر تمامی مقادیر ویژه‌ی آن مثبت باشد، آنگاه این ماتریس PD است.

[V,D] = eig(A) returns diagonal matrix D of eigenvalues and matrix V whose columns are the corresponding right eigenvectors, so that A*V = V*D.

```
%% Hessian

H = [2,1;1,2];
[V,D] = eig(H);
disp(D);
```

Command Window

```
     1     0
     0     3
```

شکل 18

در نتیجه این ماتریس PD است و تابع ما Convex می‌باشد.

سوال-4

۴) نقطه ای که مینیمم تابع در آن رخ می دهد را با صفر قرار دادن بردار گرادیان به دست آورید.

برای این کار داریم:

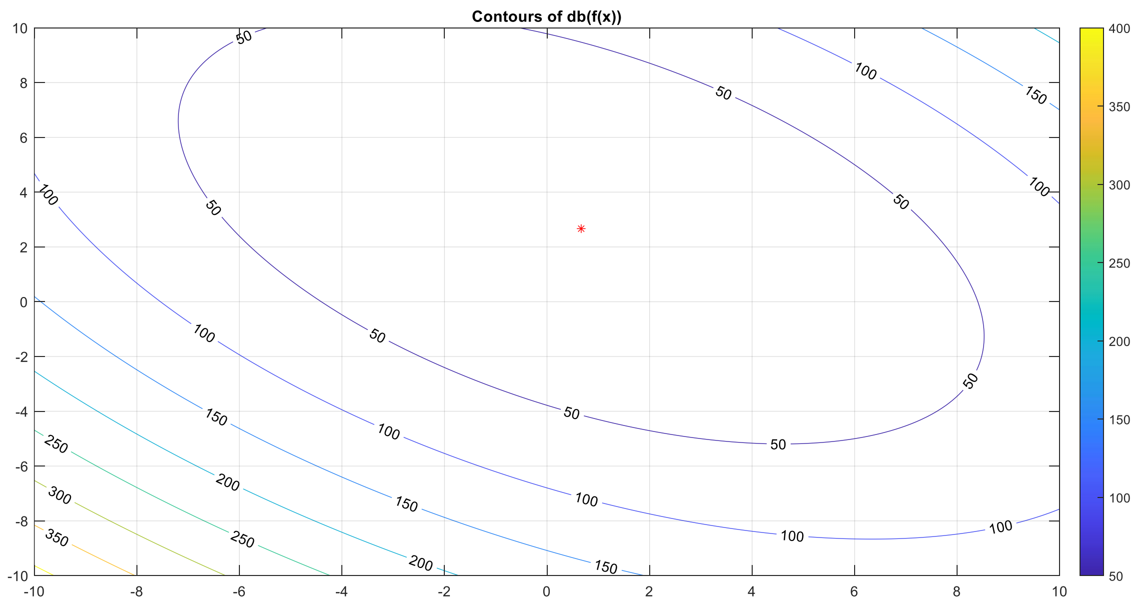
```
%% Solve g=0

A  = [2,1;1,2];
B = [4;6];

Glob_Point = A\B;

%%
```

Command Window

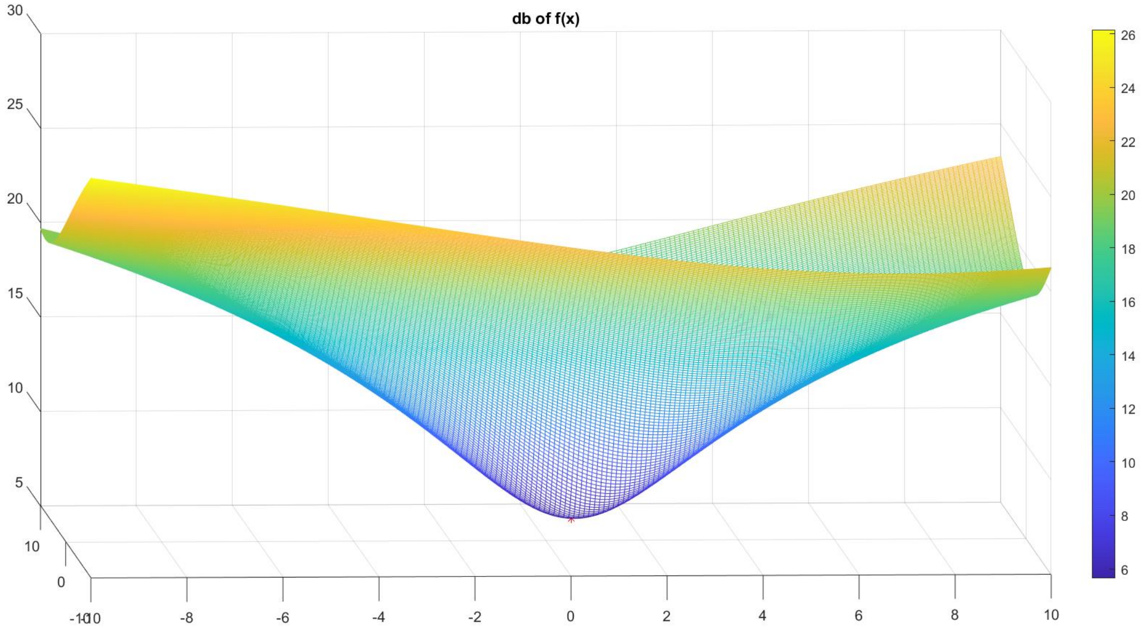
```
Glob_Point = A\B;
>> Glob_Point

Glob_Point =

    0.6667
    2.6667
```

شکل 19

و در نهایت داریم:



شکل 20



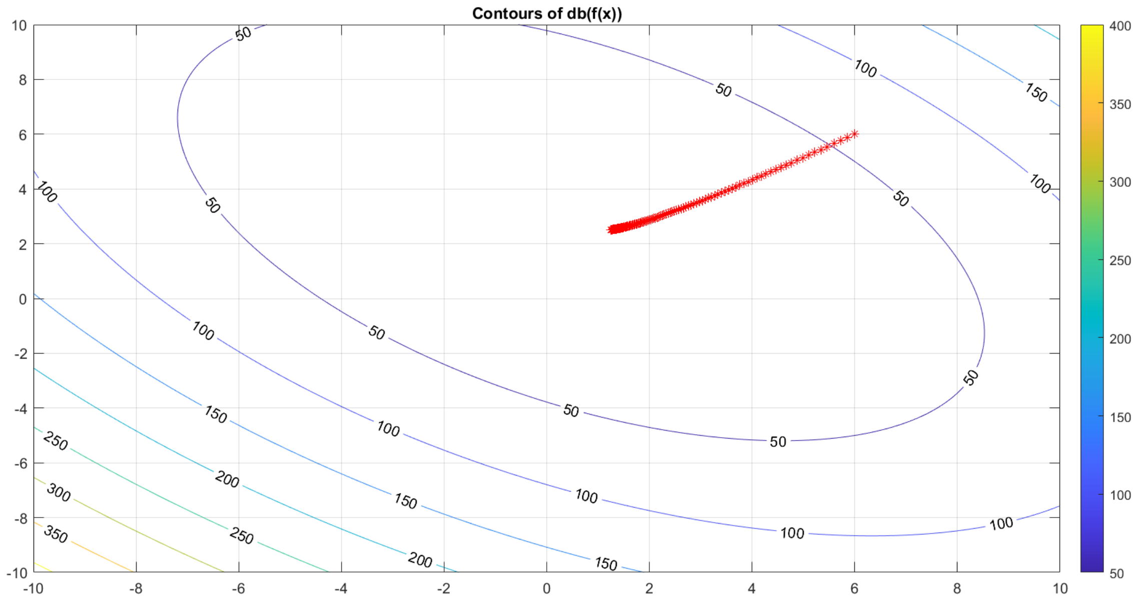
شکل 21

سوال-5:

۵) از نقطه ی اولیه $x_1 = x_2 = 6$ روش Steepest Descend را به ازای μ های 0.01 و 0.1 پیاده کنید. آیا در هر دو حالت الگوریتم همگرا شد؟ کدام یک سریع تر همگرا شد؟ نمودار همگرایی (مقدار تابع هدف بر حسب شماره iteration) را به ازای هر دو μ روی یک شکل رسم کنید.

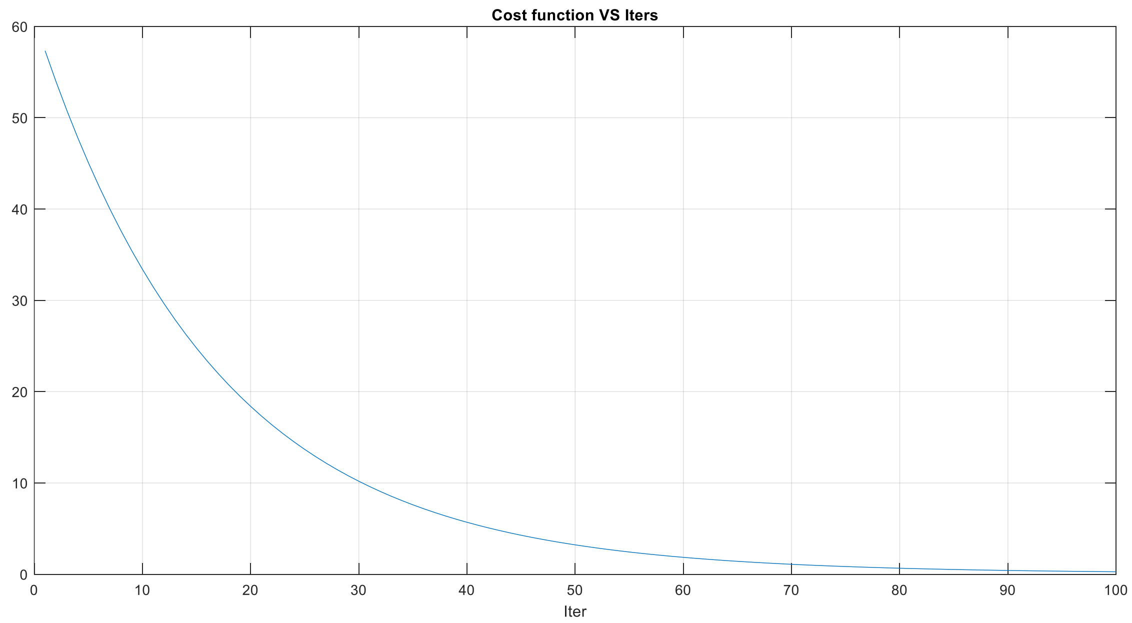
$$\mathbf{x}_1 = \mathbf{x}_0 - t_0 \nabla f(\mathbf{x}_0)$$

شکل 22



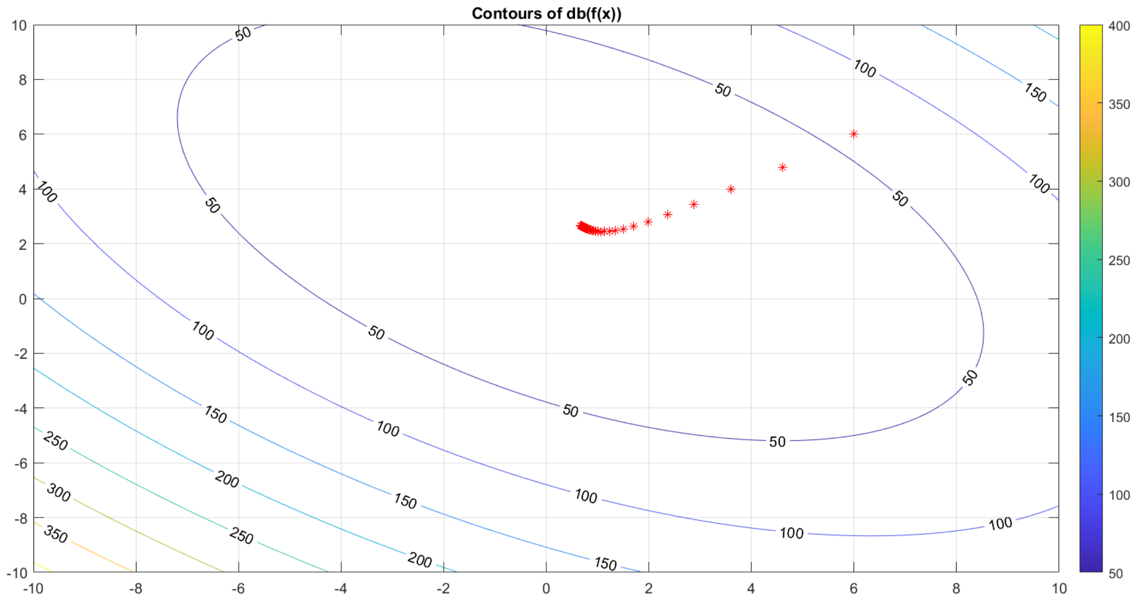
شکل 23

برای نمودار همگرایی داریم:

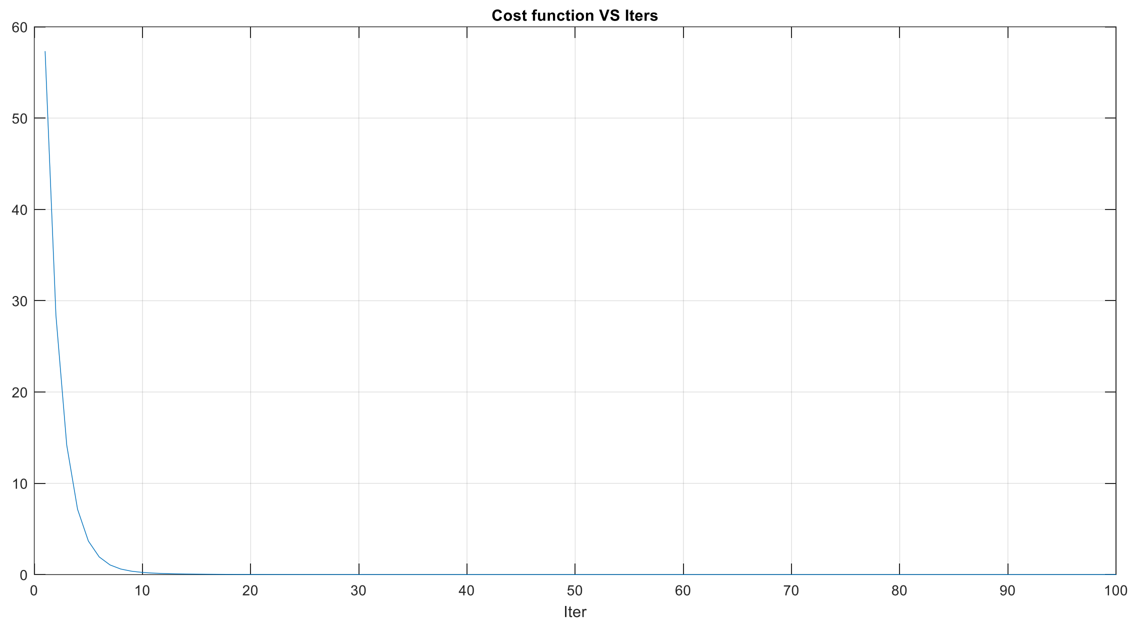


شکل 24

نمودار همگرایی برای mio=0.01



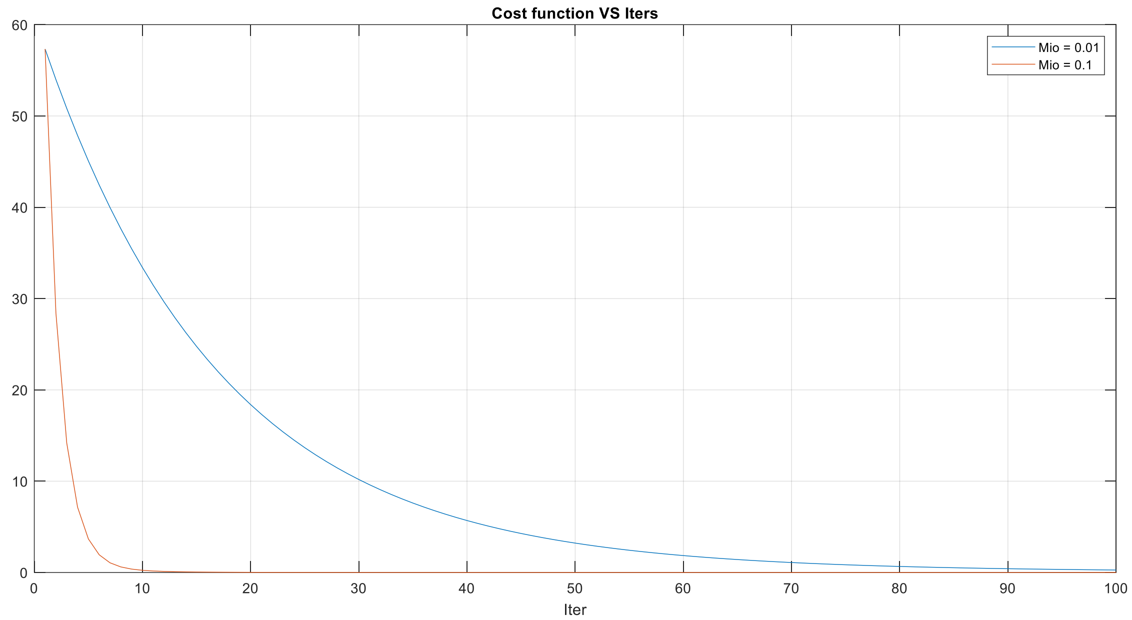
شکل 25



شکل 26

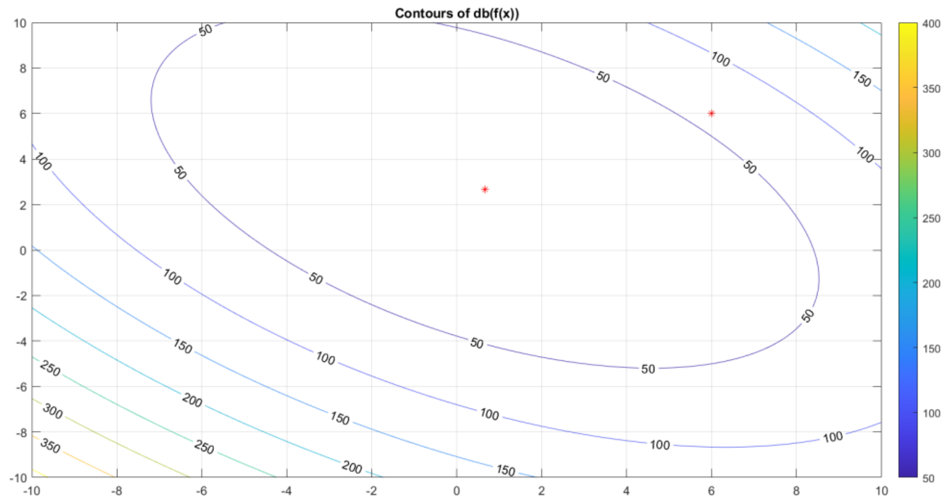
نمودار همگرایی برای mio=0. 1

واضح است که برای mio=0.1 با طول قدم های بلندتر سریع تر همگرا می‌شویم.

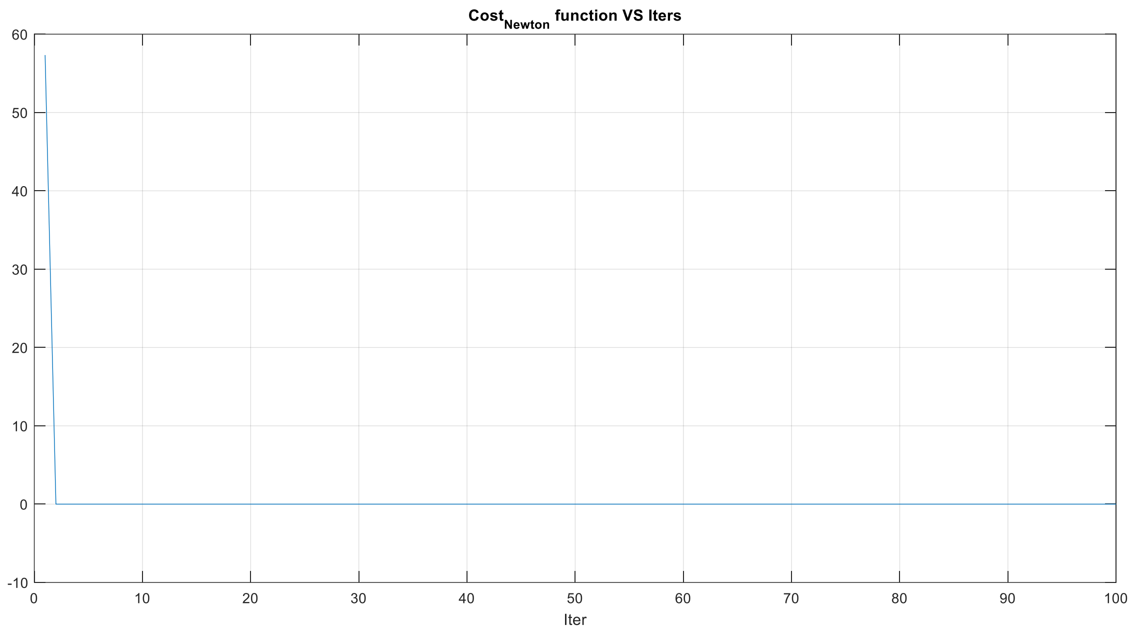


سوال-6.

۶) از نقطه ی اولیه $x_1 = x_2 = 6$ روش Newton را پیاده کنید. در چند iteration الگوریتم همگرا می شود؟ دلیل این اتفاق را شرح دهید.



شکل 27

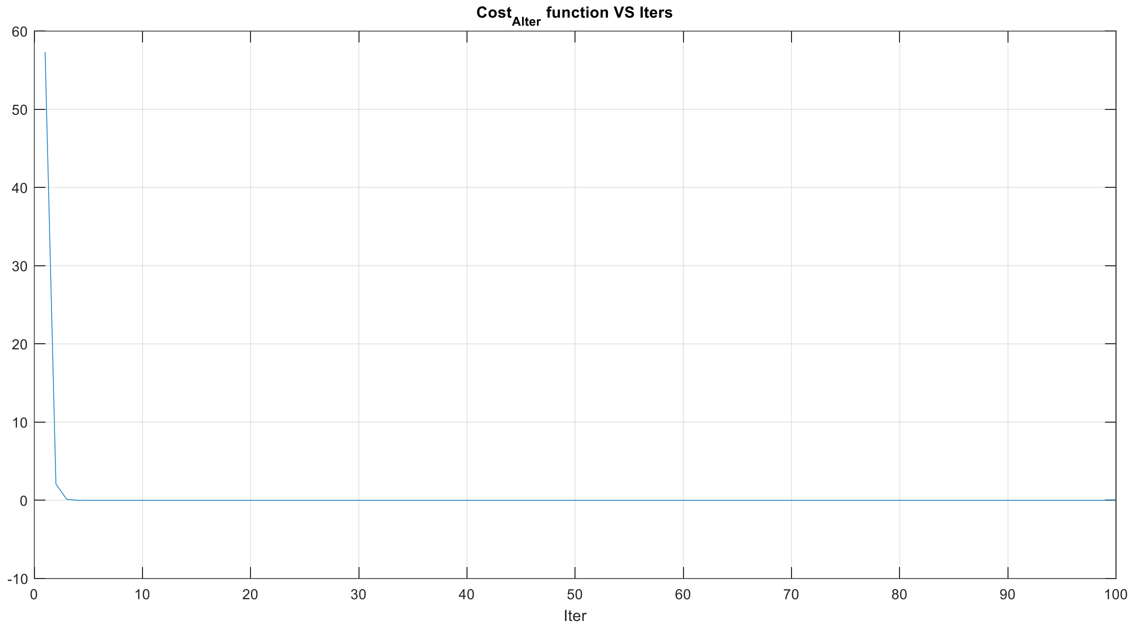


شکل 28

همانطور که مشاهده می‌شود، این روش به سرعت همگرا شده و از روش‌های قبلی سریعتر است چرا که بر اساس مشتق دوم و اطلاعات مرتبط با تحدب عمل می‌کند.

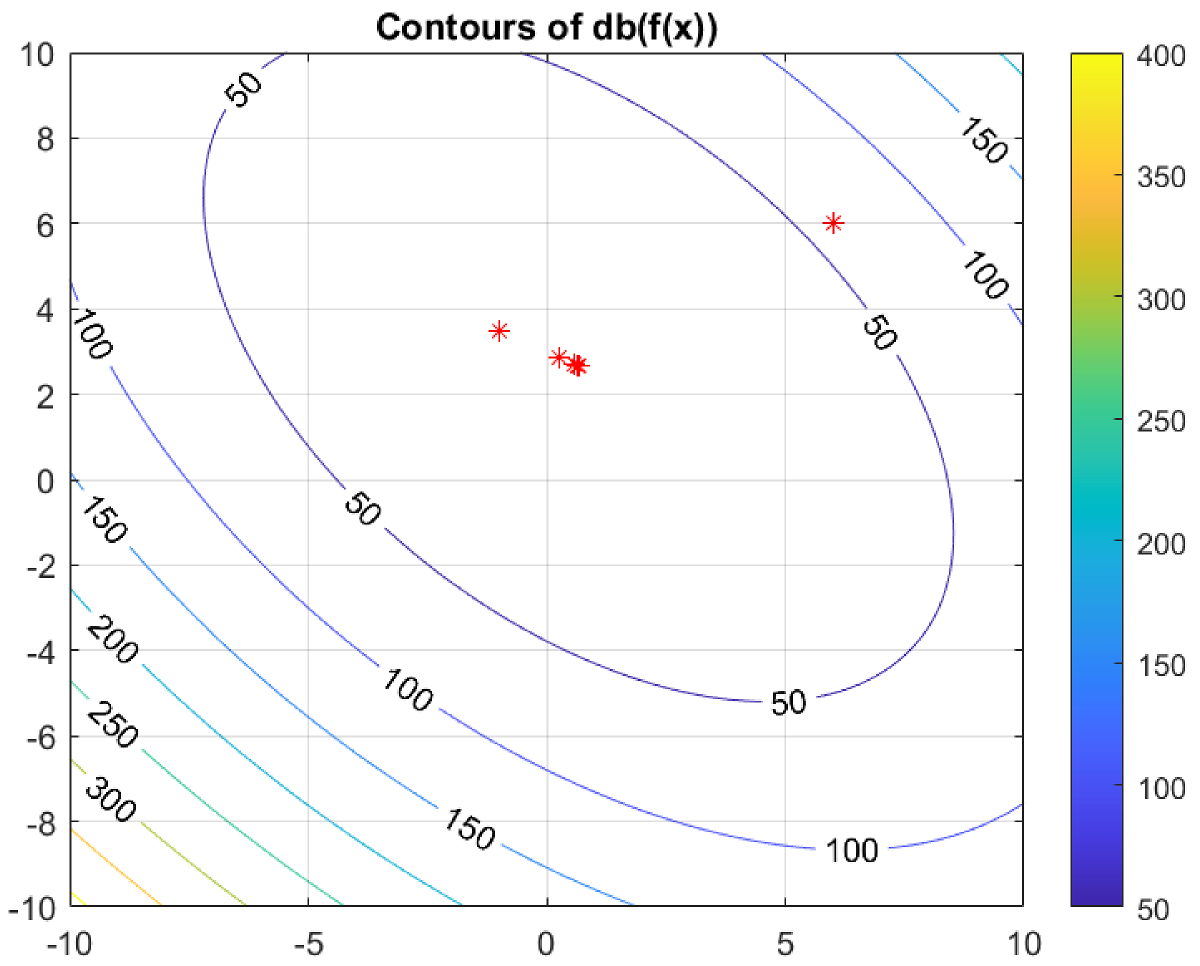
سوال-7:

۷) از نقطه ی اولیه $x_1 = x_2 = 6$ با رویکرد alternation minimization بین دو متغیر x_1 و x_2 مساله بهینه سازی را حل کنید. توجه کنید که در هر alternation، با مساله ی بهینه سازی جدیدی مواجه می شوید. چون مساله جدید ساده است، جواب فرم بسته برای آن در نظر بگیرید و دیگر احتیاجی نیست از SD یا الگوریتم های دیگر برای حل آن استفاده کنید. مسیر همگرایی را از مقدار اولیه تا مقدار بهینه در فضای دو بعدی x_1 و x_2 رسم کنید. contour ها را هم در تصویر به صورت همزمان نمایش دهید.



شکل 29

برای کانتورها داریم:

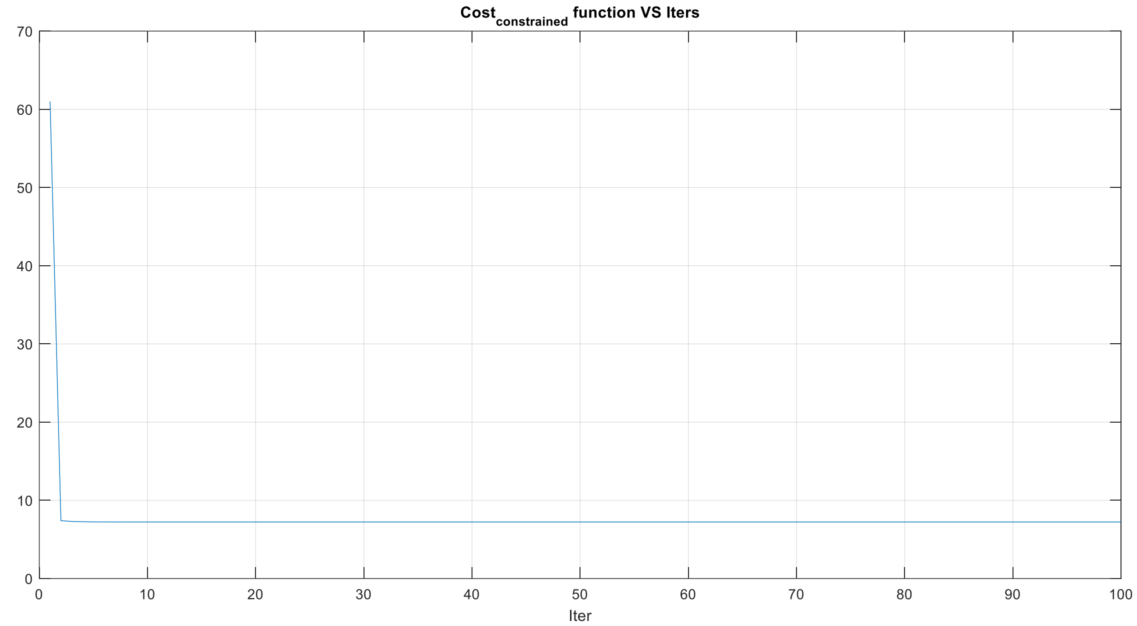


شکل 30

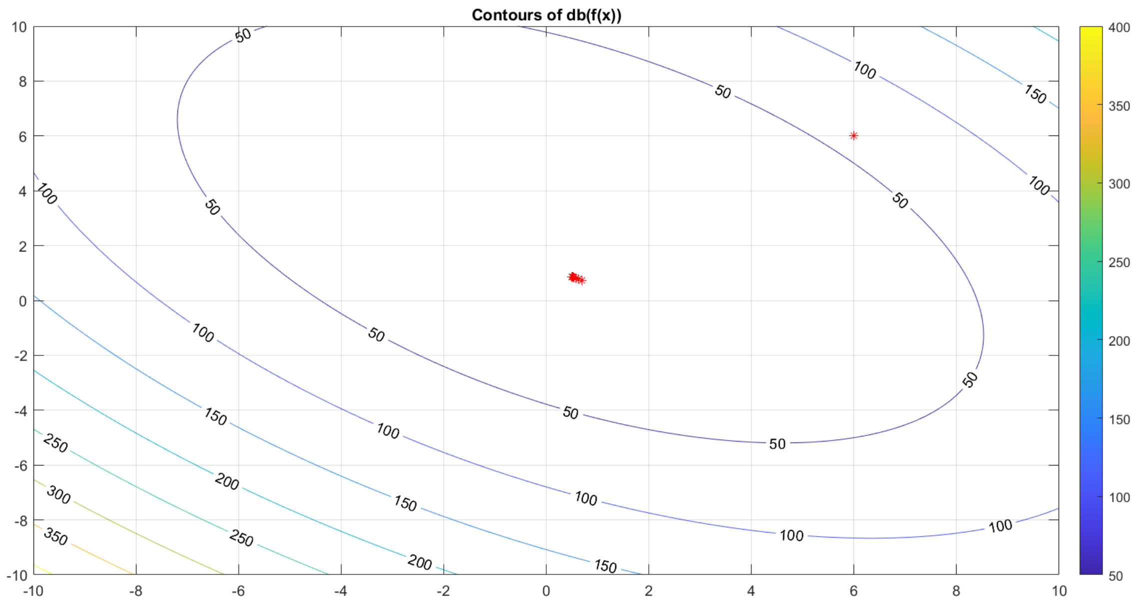
سوال-8

۸) قید $x_1^2 + x_2^2 = 1$ یا به عبارت دیگر $|\underline{x}|_2 = 1$ را به مساله اضافه می کنیم. حال تابع هدف را با رویکرد Gradient Projection از نقطه ی اولیه $x_1 = x_2 = 6$ و با استفاده از الگوریتم Steepest Descend و با $\mu = 0.1$ حل کنید. مقدار بهینه به دست آمده را گزارش کنید.

برای حل مسئله‌ی مقید، باید از روش Gradient Projection استفاده کنیم:



شکل 31



شکل 32

نقطه‌ی همگرایی:

0.4943 0.8693

که خوب با قبلی متفاوت است چرا که باید قید مسئله رعایت شود که همان واحد بودن نرم نقطه‌ی پیدا شده است.

سوال-9:

۹) با استفاده از روش ضرایب لاگرانژ، قسمت (۸) را به صورت ریاضی حل کنید و نشان دهید جواب به دست آمده در قسمت (۸) جواب مساله است.

$$\min_{x_1, x_2} f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1 - 6x_2 + 13 + x_1 x_2$$

s.t. $\|x\|_2 = 1$

$$\Rightarrow \mathcal{L}(x_1, x_2, \lambda) = x_1^2 + x_2^2 - 4x_1 - 6x_2 + 13 + x_1 x_2 - \lambda(1 - x_1^2 - x_2^2)$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial x_1} = 0 \Rightarrow 2x_1 - 4 + x_2 - \lambda(-2x_1) = 0$$

$$\frac{\partial}{\partial x_2} \mathcal{L} = 0 \Rightarrow 2x_2 - 6 + x_1 - \lambda(-2x_2) = 0$$

$$\Rightarrow \begin{cases} (2\lambda+2)x_1 + x_2 = 4 \\ x_1 + (2\lambda+2)x_2 = 6 \\ x_1^2 + x_2^2 = 1 \end{cases}$$

$$\frac{\partial}{\partial \lambda} \mathcal{L} = 0 \Rightarrow 1 - x_1^2 - x_2^2 = 0 \Rightarrow x_1^2 + x_2^2 = 1$$

$$\Rightarrow x_2 = \sqrt{1 - x_1^2} \Rightarrow \begin{cases} (2\lambda+2)x_1 + \sqrt{1 - x_1^2} = 4 \\ x_1 + (2\lambda+2)\sqrt{1 - x_1^2} = 6 \end{cases} \Rightarrow \begin{matrix} x_1 = 0.494 \\ \lambda = 2.166 \end{matrix}$$

$$\Rightarrow x_2 = 0.869 \Rightarrow f^*(x_1, x_2) = 7.236$$

با تشکر