23MAT204

MATHEMATICS FOR INTELLIGENT SYSTEMS-3

Dr. Sarada Jayan
9448455760
6364333760

How do we find x in Ax=b?

1. Use Gauss Elimination and solve Ax=b
2. Use Gauss Jacobi/Gauss Siedel Iterative methods to solve Ax=b

   Min. $\frac{1}{2}x^T A x - b^T x + c$

   Soln. $Ax = b$
3. Solve original optimization problem using steepest descent method / gradient direction method
4. Use CGM to solve original optimization problem.
5. Use GMRES to solve original optimization problem.

Direction of Descent

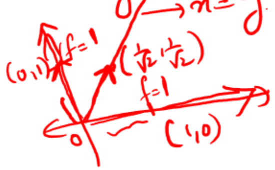
- A direction of descent at a point $x^{(k)}$ is a direction $\boldsymbol{d}$ along which the value of the function $f(\boldsymbol{x})$ decreases from $x^{(k)}$.

- Steepest descent direction:
  The direction $\boldsymbol{d}^{(k)}$ along which the function $f(\boldsymbol{x})$ decreases rapidly from the point $\boldsymbol{x}^{(k)}$ is called steepest descent direction. Steepest descent direction is along the direction of $-\nabla f\left(x^{(k)}\right)$.

Gradient of $f(x_1, \dots x_n)$ is $\nabla f(\bar{x}) = \frac{\partial f}{\partial x_1}\hat{i} + \frac{\partial f}{\partial x_2}\hat{j} + \dots$

Application of grad $\rightarrow$ ∇f gives the dirn. of max. increase of f.

$f = x + y$



$\bar{d} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \nabla f_{(1,1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Steepest Descent Algorithm

- The iterative formula for method of steepest descent is

  $x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)}$

- The descent direction $d^{(k)}$ is along the steepest descent direction, $d^{(k)} = -\nabla f\left(x^{(k)}\right)$
- The step length $\alpha^{(k)}$ is obtained by performing a unidirectional search from $x^{(k)}$ along the direction $d^{(k)}$ i.e., by minimizing,

  $\varphi\left(\alpha^{(k)}\right) = f(x^{(k)} + \alpha^{(k)} d^{(k)})$

  $\left(\varphi'\left(\alpha^{(k)}\right) = 0, \ \varphi''\left(\alpha^{(k)}\right) > 0 \right)$

<u>Algorithm:</u>

Step 1: Chose an initial starting point $x^{(0)}$ and a termination parameter ε.

Step 2: Compute steepest descent, $d^{(0)} = -\nabla f\left(x^{(0)}\right)$

Step 3: Compute the step length, $\alpha^{(0)}$[using unidirectional search from $x^{(0)}$ along $d^{(0)}$].

Step 4: Evaluate $x^{(1)} = x^{(0)} + \alpha^{(0)} d^{(0)}$

Step 5: Compute $\left\|\nabla f\left(x^{(1)}\right)\right\|$. If $\left\|\nabla f\left(x^{(1)}\right)\right\| < \varepsilon$, stop and mention $x^{(1)}$ is minimum,

Else go to step 2 with $x^{(0)} = x^{(1)}$.

[While manually solving these problems gradients in step 2 can be directly found. But while coding the numerical formula to evaluate gradients need to be used.]

Exercise

1. Find a minimum for the function, $f(x,y) = (x-1)^2 + (y-2)^2$ starting from the point (10,-1), using steepest descent method. Choose the termination parameter $\varepsilon = 0.1$

Exercise

1. Find a minimum for the function, $f(x,y) = (x-1)^2 + (y-2)^2$ starting from the point (10,-1), using steepest descent method. Choose the termination parameter $\varepsilon = 0.1$

$f(x,y) = (x-1)^2 + (y-2)^2, \quad \bar{x}^{(0)} = (10,-1), \quad \varepsilon = 0.1$

$\nabla f = [2(x-1), 2(y-2)]$

$\nabla f(10,-1) = [18, -6]$

$\bar{d}^{(0)} = -\nabla f(\bar{x}^{(0)}) = [-18, 6]$

Unidirectional search from $\bar{x}^{(0)}$ along $\bar{d}^{(0)}$:

$s(\alpha) = \bar{x}^{(0)} + \alpha \bar{d}^{(0)} = (10,-1) + \alpha(-18,6)$

$= [10-18\alpha, -1+6\alpha]$

$f(s(\alpha)) = [10-18\alpha-1]^2 + [-1+6\alpha-2]^2$

$\phi(\alpha) = (9-18\alpha)^2 + (6\alpha-3)^2$

$\phi(\alpha) = (9-18\alpha)^2 + (6\alpha-3)^2$

Minimum of $\phi(\alpha) \rightarrow \phi'(\alpha) = 0$

$\Rightarrow 2(9-18\alpha)(-18) + 2(6\alpha-3)(6) =$

$\Rightarrow -324 + 648\alpha + 72\alpha - 36 = 0$

$\Rightarrow 720\alpha - 360 = 0$

$\Rightarrow \alpha^* = \frac{360}{720} = \frac{1}{2}$

$\bar{x}_1 = \bar{s}(\alpha^*) = [1, 2]$ [Or $\bar{x}_1 = \bar{x}^{(0)} + \alpha^{(0)} \bar{d}^{(0)}$]

$= (10,-1) + \frac{1}{2}(-18,6)$

$\nabla f(\bar{x}_1) = \nabla f(1,2) = [0, 0]$.

$\Rightarrow \|\nabla f(1,2)\| = 0 < \varepsilon$

$\therefore$ Min. of f is at (1,2).

Exercise

2. Apply three iterations of steepest descent method to find a minimum for the function, $f(x,y) = 2x^2 - 2xy + y^2$ starting from the point (1,2).

Exercise

2. Apply three iterations of steepest descent method to find a minimum for the function, $f(x,y) = 2x^2 - 2xy + y^2$ starting from the point (1,2).

$\nabla f = \begin{pmatrix} 4x - 2y \\ -2x + 2y \end{pmatrix}$

$x^{(0)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $d^{(0)} = -\nabla f\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$ *

$x^{(1)} = x^{(0)} + \alpha d^{(0)} = \begin{pmatrix} 1 \\ 2 - 2\alpha \end{pmatrix}$

Unidirectional search.

$f(x^{(1)}(\alpha)) = f(\alpha) = 2\cdot 1 - 2\cdot 1(2 - 2\alpha) + (2 - 2\alpha)^2$

$f'(\alpha) = 0 \Rightarrow 4 + 2(2 - 2\alpha)(-2) = 0 \Rightarrow 4 - 8 + 8\alpha = 0$ } $\alpha^* = 1/2$ $f''(\alpha) = 8 > 0$ $\Rightarrow \alpha^* = 1/2$ gives minimum

$\therefore x^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

2nd iteration

$x^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $d^{(1)} = -\nabla f(x^{(1)}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$ *

$x^{(2)} = x^{(1)} + \alpha d^{(1)} = \begin{pmatrix} 1 - 2\alpha \\ 1 \end{pmatrix}$

Unidir. Search $f(\alpha) = 2(1 - 2\alpha)^2 - 2(1 - 2\alpha) + 1$

$f'(\alpha) = 0 \Rightarrow 2\cdot 2(1 - 2\alpha)(-2) + 4 = 0$

$\Rightarrow -8 + 16\alpha + 4 = 0 \Rightarrow \alpha = 1/4$

$\boxed{\begin{matrix} d^{(0)} = \begin{pmatrix} 0 \\ -2 \end{pmatrix} \\ d^{(1)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \\ d^{(2)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \end{matrix}}$

orthogonal

$\alpha = 1/4$, $f''(\alpha) = 16 > 0$

$\Rightarrow \alpha^* = 1/4$ gives minimum.

$x^{(2)} = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$, $\nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

3rd iteration

$x^{(2)} = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$, $d^{(2)} = -\nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ *

$x^{(3)} = x^{(2)} + \alpha d^{(2)} = \begin{pmatrix} \frac{1}{2} \\ 1 - \alpha \end{pmatrix}$

Unid. $f(\alpha) = 2\left(\frac{1}{2}\right)^2 - 2\cdot\frac{1}{2}(1 - \alpha) + (1 - \alpha)^2$

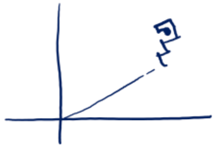
$f'(\alpha) = 0 \Rightarrow 1 + 2(1 - \alpha)(-1) = 0$

$1 - 2 + 2\alpha = 0$

$\Rightarrow \alpha = 1/2$

$f''(\alpha) = 2 > 0 \Rightarrow \alpha^* = 1/2$ is min. of $f(\alpha)$

$x^{(3)} = x^{(2)} + \alpha d^{(2)} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$



Exercise

3. Apply 4 iterations of steepest descent method to find the minimum of the function, $f(x_1, x_2) = 3x_1^2 - 4\, x_1 x_2 + 2x_2^2 + 4\, x_1 + 6$ starting from the origin. [Evaluate the gradients in each iteration analytically].

Ans:

k	$\mathbf{X^{(k)}}$	$\nabla f(\mathbf{X^{(k)}})$	$\mathbf{d^{(k)}}$	α_k
1	(0,0)	(4,0)	(-4,0)	1/6
2	(-2/3,0)	(0,8/3)	(0,-8/3)	1/4
3	(-2/3,-2/3)	(8/3,0)	(-8/3,0)	1/6
4	(-10/9,-2/3)	(0,16/9)	(0,-16/9)	1/4

Steepest Descent method

- This method is also called as the gradient descent method as it uses the negative gradient as the search direction in each iteration.
- **The method produces successive directions that are perpendicular to each other.**
- When the point is away from the optimum, the method makes good progress towards the optimum.
- Near the optimum due to zigzagging, the convergence becomes very slow.

Exercise

4. Consider the $f(x,y) = \sin x + 4x^2 + y^3 - 3y + 2$.

(a)Is (1,1) or (2,-1) a descent direction from (0,0)?

(b)Find the steepest descent direction for $f(x,y)$ from (0,0).

(a) $\nabla f = \begin{pmatrix} \cos x + 8x \\ 3y^2 - 3 \end{pmatrix}$ (2m) ; $\nabla f(0,0) = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ (1m)

(i) $\nabla f \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -2 < 0$. (1m)

$\therefore$ $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a descent dirn.

(ii) $\nabla f \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} = 5 > 0$ (1m)

$\therefore$ (2,-1) is not a descent dirn

(b) $\cancel{\nabla f(0,0)} =$ Steepest descent dirn $= \begin{pmatrix} -1 \\ 3 \end{pmatrix}$.

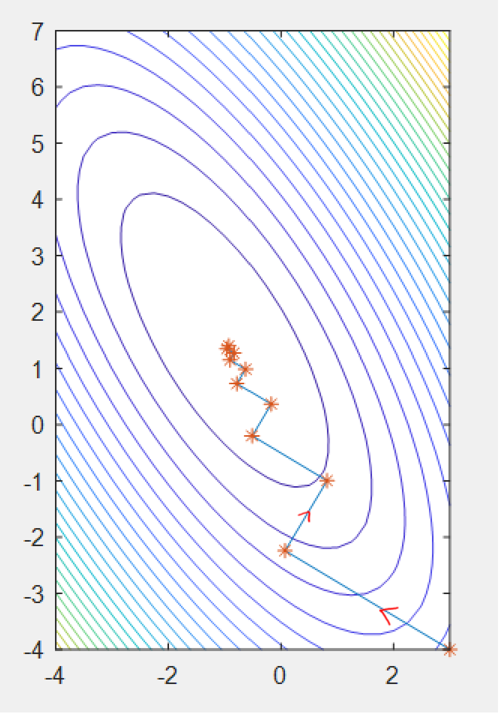
5. Find the steepest descent direction for the function, $f(x,y) = 5x^3 + e^{-3y}$ from (1,1).

Ans: Steepest Descent = - ∇f

$$\nabla f = \begin{pmatrix} 15x^2 \\ -3e^{-3y} \end{pmatrix}$$

Steepest Descent = - $\nabla f \; at \; (1,1) = \begin{pmatrix} 15 \\ -3e^{-3} \end{pmatrix}$

Solve original optimization problem using gradient direction. It is an iterative method



$$Let\ \ f(x) = \frac{1}{2}x^T A x - b^T x + c,\ \ x \in R^n, A = A^T, c \in R$$

$Start$ with arbitrary x_0

$\nabla f(x) = Ax - b \Rightarrow\ Gradient\ at\ x = x_0$ is $Ax_0 - b$

We denote this as $g(x_0)$

$-g(x_0)$ steepest descent dirn.

$Negative$ gradient is $r_0 = b - Ax_0$. r_0 is called residual

New updated x is $x_1 = x_0 - \alpha\, g(x_0)$

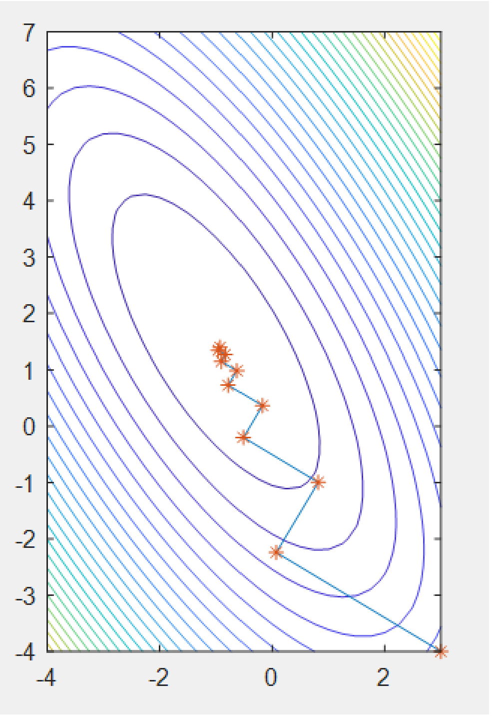
$What\ is$ a good α ?

The one obtained using unidirectional search.

f P $\nabla f \rightarrow$ fn. increases fastly

$f(x_1, x_2, \ldots x_n)$.
$\nabla f(x^{(k)})$
Steepest descent dirn. $= -\nabla f(x^{(k)})$
$d^{(k)} = -\nabla f(x^{(k)})$
$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$
or
$= x^{(k)} - \alpha_k \underbrace{g^{(k)}}_{\nabla f(x^{(k)})}$

Solve original optimization problem using gradient direction. It is an iterative method



$Let\ \ f(x) = \dfrac{1}{2}x^T A x - b^T x + c,\ \ x \in R^n, A = A^T, c \in R$

$Start$ with arbitrary x_0

$\nabla f(x) = Ax - b \Rightarrow\ Gradient\ at\ x = x_0$ is $Ax_0 - b$

We denote this as $g(x_0)$

$Negative$ gradient is $r_0 = b - Ax_0$. r_0 is called residual

New updated $x\ is\ \ x_1 = x_0 - \alpha\, g(x_0)$

$What\ is$ a good α ?

The one obtained using unidirectional search.

Solution

$Let\ \ f(x) = \frac{1}{2} x^T A x - b^T x + c,\ \ x \in R^n, A = A^T,\ \ \ c \in R$

$Start$ with arbitrary x_0 (position vector)

$\nabla f(x) = Ax - b \Rightarrow\ Gradient\ at\ x = x_0$ is $g(x_0) = Ax_0 - b$

$Note$: Never forget that g(.) is a vector (displacement vector)

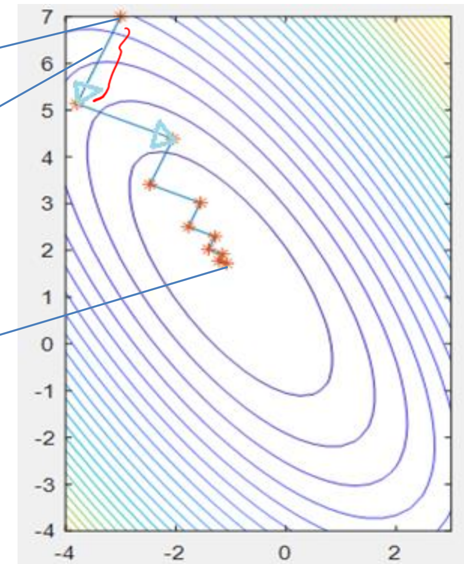
$Let\ x_1 = x_0 - \alpha\ g(x_0).\ \ We\ donot\ know\ \alpha\ to\ compute\ x_1$

Let us find it through optimization.

$x_0 = \begin{pmatrix} -3 \\ 7 \end{pmatrix}$

$-\alpha g(x_0) = \begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$

$x^* = \begin{pmatrix} -1 \\ 3/2 \end{pmatrix}$



$$f(x_0 - \alpha\ g(x_0)) = \underbrace{\frac{1}{2}(x_0 - \alpha\ g(x_0))^T A(x_0 - \alpha\ g(x_0))}_{\text{scalar fn.}} - b^T (x_0 - \alpha\ g(x_0)) + c$$

$$= -\alpha x_0^T A g(x_0) + \frac{\alpha^2}{2}(g(x_0))^T A g(x_0) + \alpha b^T\ g(x_0) + K,$$ where K is constant

$$\frac{df}{d\alpha} = -x_0^T A g(x_0) + b^T\ g(x_0) + \alpha\ (g(x_0))^T A g(x_0)$$

$$\frac{df}{d\alpha} \Rightarrow 0 \Rightarrow \alpha = \frac{(g(x_0))^T (Ax_0) - (g(x_0))^T b}{(g(x_0))^T A g(x_0)} = \frac{(g(x_0))^T (\overbrace{Ax_0 - b}^{g(x_0)})}{(g(x_0))^T A g(x_0)} = \frac{(g(x_0))^T g(x_0)}{(g(x_0))^T A g(x_0)}$$

$g(x_0) = Ax_0 - b$

$\alpha_0 = \frac{g(x_0)^T g(x_0)}{g(x_0)^T A g(x_0)}$

$x_1 = x_0 - \alpha_0 g(x_0)$

$g(x_1) = Ax_1 - b$

$\alpha_1 = \frac{g(x_1)^T g(x_1)}{g(x_1)^T A g(x_1)}$.

$x_2 = x_1 - \alpha_0 g(x_1)$

Solution

$Let\ \ f(x) = \frac{1}{2} x^T A x - b^T x + c, \ \ x \in R^n, A = A^T, \ \ \ c \in R$

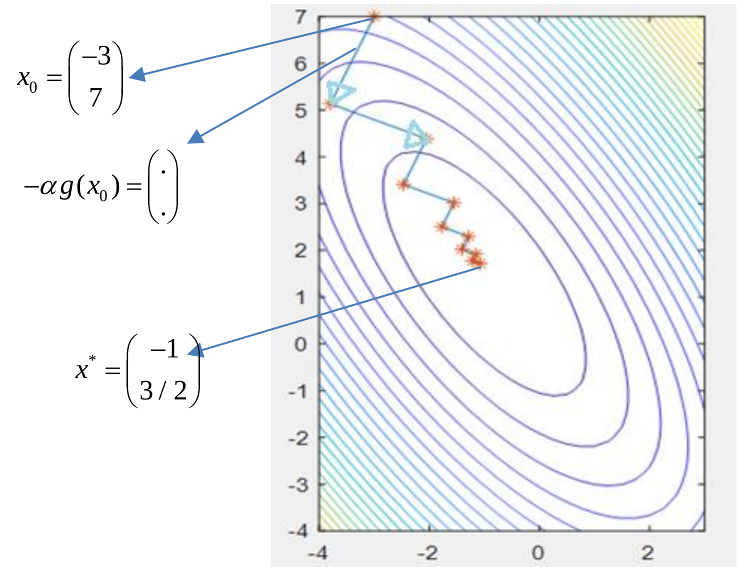
$Start$ with arbitrary x_0 (position vector)

$\nabla f(x) = Ax - b \Rightarrow \ Gradient\ at\ x = x_0$ is $g(x_0) = Ax_0 - b$

$Note:$ Never forget that g(.) is a vector (displacement vector)

$Let\ x_1 = x_0 - \alpha\ g(x_0).\ \ We\ donot\ know\ \alpha\ to\ compute\ x_1$

Let us find it through optimization.



$$f\left(x_0 - \alpha\ g(x_0)\right) \ = \frac{1}{2}\left(x_0 - \alpha\ g(x_0)\right)^T A\left(x_0 - \alpha\ g(x_0)\right) - b^T\left(x_0 - \alpha\ g(x_0)\right) + c$$

$$= -\alpha x_0{}^T A g(x_0) + \frac{\alpha^2}{2}\left(g(x_0)\right)^T A g(x_0) + \alpha b^T\ g(x_0) + K, \quad \text{where K is constant}$$

$$\frac{df}{d\alpha} = -x_0{}^T A g(x_0) + b^T\ g(x_0) + \alpha\ \left(g(x_0)\right)^T A g(x_0)$$

$$\frac{df}{d\alpha} \Rightarrow 0 \Rightarrow \alpha = \frac{\left(g(x_0)\right)^T\left(Ax_0\right) - \left(g(x_0)\right)^T b}{\left(g(x_0)\right)^T A g(x_0)} = \frac{\left(g(x_0)\right)^T\left(Ax_0 - b\right)}{\left(g(x_0)\right)^T A g(x_0)} = \frac{\left(g(x_0)\right)^T g(x_0)}{\left(g(x_0)\right)^T A g(x_0)}$$

1. Solve the system: $2x_1-x_2=1$; $-x_1+2x_2=0$; using gradient descent algorithm with initial vector as $(0,0)^T$

→ steepest descent direction

$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$; $\bar{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; $\bar{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$; $\boxed{g(x) = Ax - b}$

$\boxed{d^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}}$

Iteration 1:-

$g(x^{(0)}) = Ax^{(0)} - b = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

$\alpha_0 = \dfrac{\langle g(x^{(0)}), g(x^{(0)}) \rangle}{\langle Ag(x^{(0)}), g(x^{(0)}) \rangle} = \dfrac{1}{\begin{pmatrix} -2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix}} = \dfrac{1}{2}$

$x^{(1)} = x^{(0)} - \alpha_0 g(x^{(0)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$

$g(x^{(1)}) = Ax^{(1)} - b = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}$, $\|g(x^{(1)})\| \neq 0$

Iteration 2

$\alpha_1 = \dfrac{g(x^{(1)})^T g(x^{(1)})}{g(x^{(1)})^T A g(x^{(1)})} = \dfrac{1/4}{1/2} = 1/2 = 0.5$

$\boxed{d^{(1)} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}}$

$x^{(2)} = x^{(1)} - \alpha_1 g(x^{(1)}) = \begin{pmatrix} 0.5 \\ 0.25 \end{pmatrix}$

$g(x^{(2)}) = Ax^{(2)} - b = \begin{pmatrix} -0.25 \\ 0 \end{pmatrix}$, $\|g(x^{(2)})\| \neq 0$

$\boxed{d^{(2)} = \begin{pmatrix} 0.25 \\ 0 \end{pmatrix}}$

Iteration 3

$\alpha_2 = \dfrac{g(x^{(2)})^T g(x^{(2)})}{g(x^{(2)})^T A g(x^{(2)})} = 0.5$

$x^{(3)} = x^{(2)} - \alpha_2 g(x^{(2)}) = \begin{pmatrix} 5/8 \\ 1/4 \end{pmatrix} = \begin{pmatrix} 0.625 \\ 0.25 \end{pmatrix}$

$g(x^{0})^T g(x^{(0)}) = \langle g(x^{0}), g(x^{(0)}) \rangle$

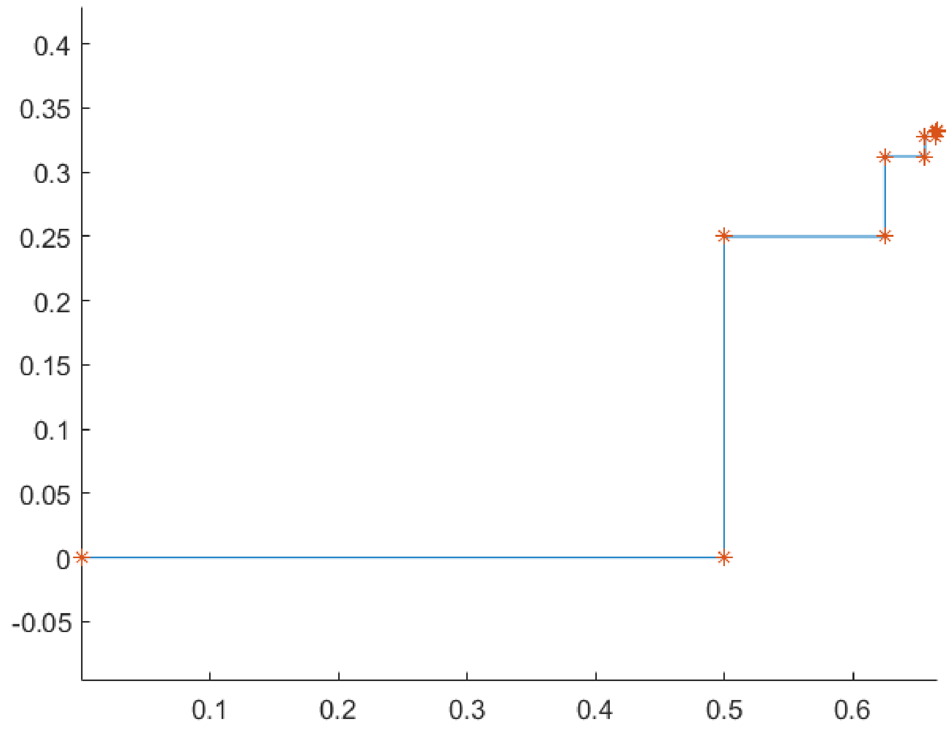
$g(x^{(0)})^T A g(x^{(0)}) = \langle g(x^{(0)}), Ag(x^{(0)}) \rangle$
$= \langle Ag(x^{0}), g(x^{0}) \rangle$

Direction of descent selected in each iteration is $\perp$ to the direction in the previous iteration.

After 3 iterations the soln we get is $\bar{x} = \begin{pmatrix} 0.625 \\ 0.25 \end{pmatrix}$

```
A=[2,-1;-1,2]; b=[1;0];
xo=[0;0]; % starting point, column vector
xa=xo; % store in array
for i=1:10
g=A*xo-b;
alpha=g'*g/(g'*A*g);
xn=xo-alpha*g;
xa=[xa xn];
xo=xn;
errnorm=norm(A*xn-b,2);
if errnorm <0.001
break;
end
end
hold on
plot(xa(1,:),xa(2,:))
hold on
plot(xa(1,:),xa(2,:),'*')
axis equal
xa
```

g(xn)




```
xa = 2×11
     0    0.5000    0.5000    0.6250    0.6250    0.6563    0.6563    0.6641    0.6641    0.6660    0.6660
     0         0    0.2500    0.2500    0.3125    0.3125    0.3281    0.3281    0.3320    0.3320    0.3330
```

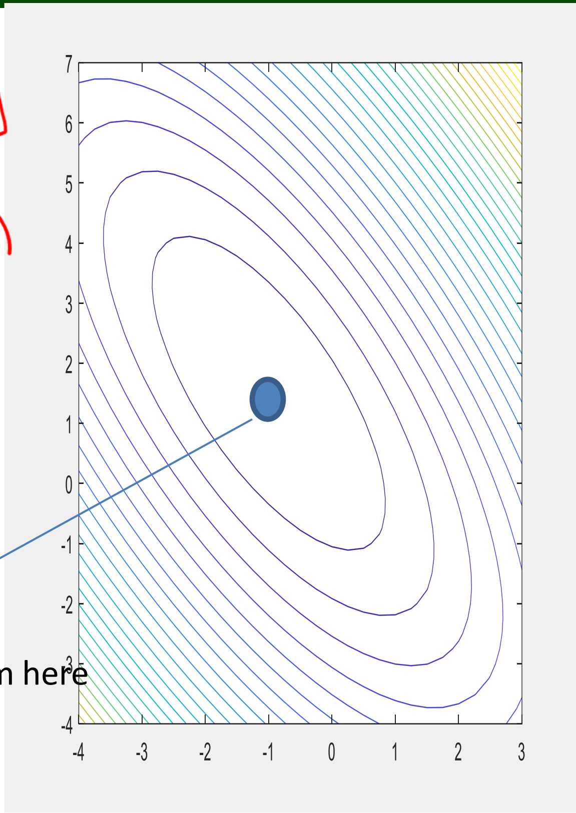
Find Minimizer of

$$f(x_1,x_2) = 2x_1^2 + 1x_2^2 + 2x_1x_2 + x_1 - x_2 = \frac{1}{2}\left[\underbrace{4x_1^2 + 2x_2^2 + 4x_1x_2}_{x^TAx}\right] + \underbrace{x_1 - x_2}_{-b^Tx}$$

$$= \frac{1}{2}\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}; \quad b = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$= \frac{1}{2}x^T Ax - b^T x$$



f(x) is Minimum here

OR

Solve

$$\left.\begin{aligned} 4x_1 + 2x_2 &= -1 \\ 2x_1 + 2x_2 &= 1 \end{aligned}\right\} \quad \begin{aligned} 2x_1 &= -2 \\ -4 + 2x_2 &= -1; \end{aligned} \quad \boxed{\begin{aligned} x_1 &= -1 \\ x_2 &= 3/2 \end{aligned}}$$

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

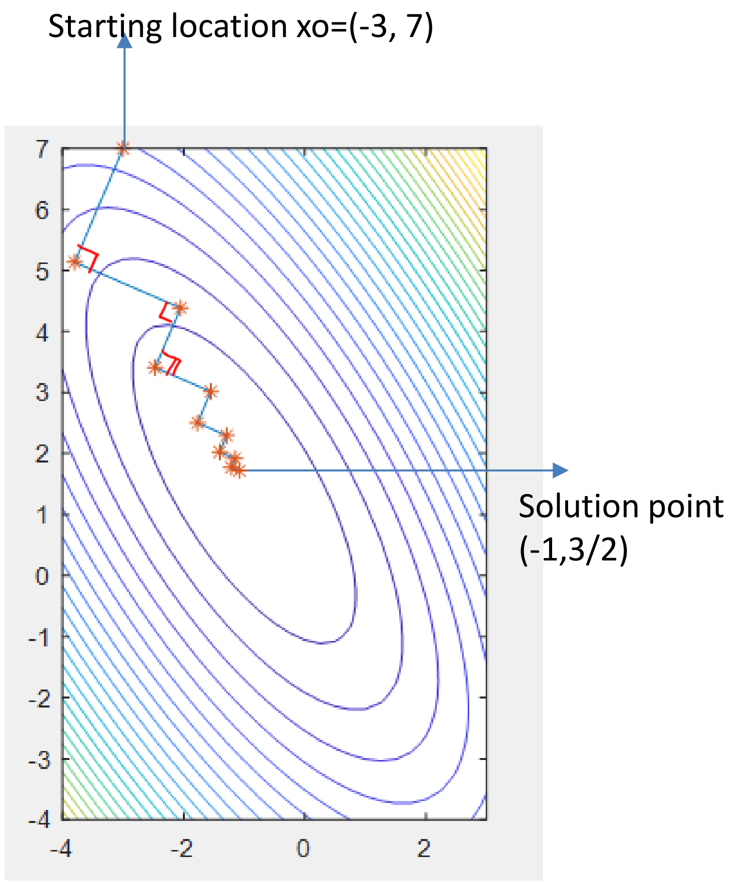
$$\begin{pmatrix} \partial f/\partial x \\ \partial f/\partial y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 4x_1 + 2x_2 + 1 \\ 2x_2 + 2x_1 - 1 \end{pmatrix}$$

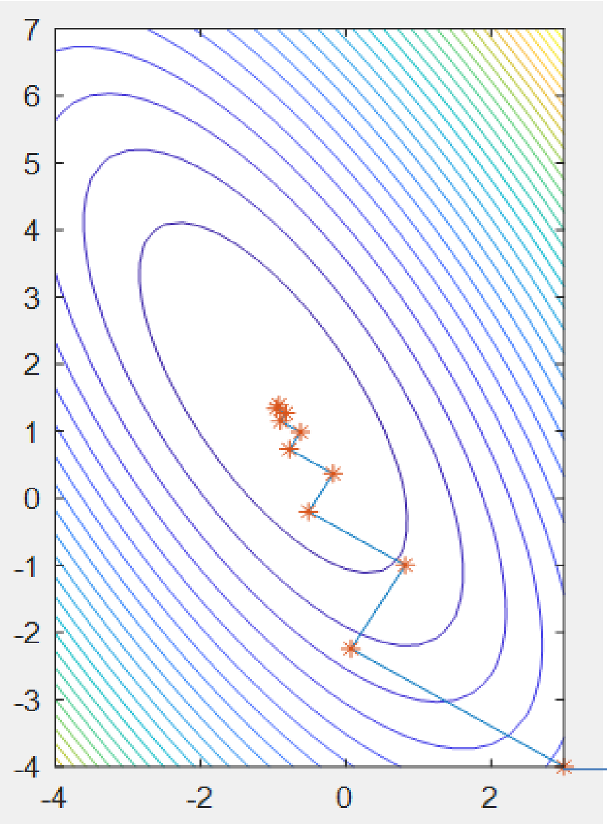
```
x=-4:.25:3;
y=-4:.25:7;
[X1,X2] = meshgrid(x,y);
Z = 2*X1.^2+X2.^2+2*X1.*X2+X1-X2;
contour(X1,X2,Z,30);
```

The solution is x*=(-1,1.5)

```
x1=-4:.25:3;
x2=-4:.25:7;
[X1,X2] = meshgrid(x1,x2);
Z = 2*X1.^2+X2.^2+2*X1.*X2+X1-X2;
contour(X1,X2,Z,30);
xo=[-3;7]; % starting point, column vector
A=[4 2; 2 2]; b=[-1;1];
xa=xo; % store in array
for i=1:10
    g=A*xo-b;
    alpha=g'*g/(g'*A*g);
    xn=xo-alpha*g;
    xa=[xa xn];
    xo=xn;
    errnorm=norm(A*xn-b,2);
    if errnorm <0.001
        break;
    end
end
hold on
 plot(xa(1,:),xa(2,:))
 hold on
 plot(xa(1,:),xa(2,:),'*')
 axis equal
```

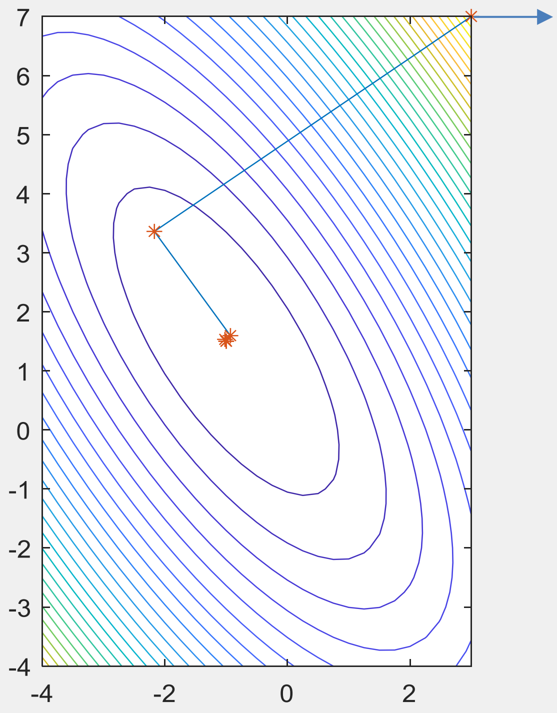


Starting point (3,-4)



Starting location xo=(3, -4)

Starting point (3,7)

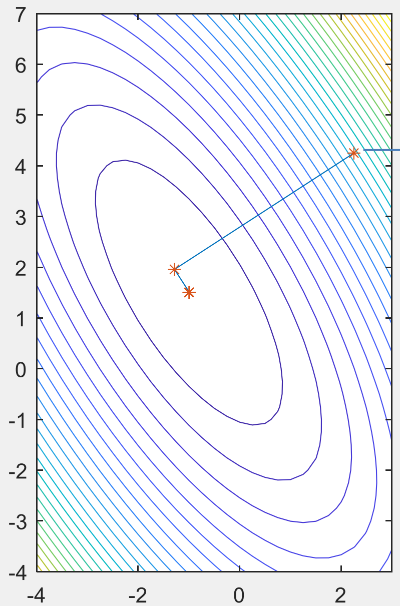


Starting location xo=(3, 7)

```
isovalues = (-3.0:0.5:3.0);
contour(X,Y,Z,isovalues);
```

https://www.bu.edu/tech/support/research/training-consulting/online-tutorials/visualization-with-matlab/

Starting point (2.25 4.25)



Starting location xo=(2.25, 4.25)

Observation

1. Number of steps required to converge for the gradient descent method depends on the starting point.
2. There is no way to predict the number of steps required.

If x is from R^n , can we get convergence in n steps?.

Yes, we have to use a special direction called Conjugate direction.