Date. Page.

# HW 8

1. a) $\nabla f(x)^T d = \nabla f(x)^T [-(\nabla f(x))_j e_j] = -\|\nabla f(x)\|_\infty^2 \le 0$.
d is descent direction

b) $\nabla f(x)^T d = \nabla f(x)^T [-\frac{\nabla f(x)}{\|\nabla f(x)\|}] = -\frac{\|\nabla f(x)\|^2}{\|\nabla f(x)\|} \Leftarrow = -\|\nabla f(x)\| < 0$
d is descent direction

c) No division by 0 since $\max \nabla^2 f(x)_{ii}, \epsilon \ge 0$., $\nabla f(x)^T d = -\sum_{i=1}^{n} \frac{(\nabla f(x)_i)^2}{\max(\nabla^2 f(x)_{ii}, \epsilon)} \le 0$
→ descent property relies on $(\nabla^2 f(x))_{ii}$.
→ directional derivation in the direction decrease $f(x)$ locally.
given $\nabla f(x) \neq 0$ → $d$ is well-defined and descent direction

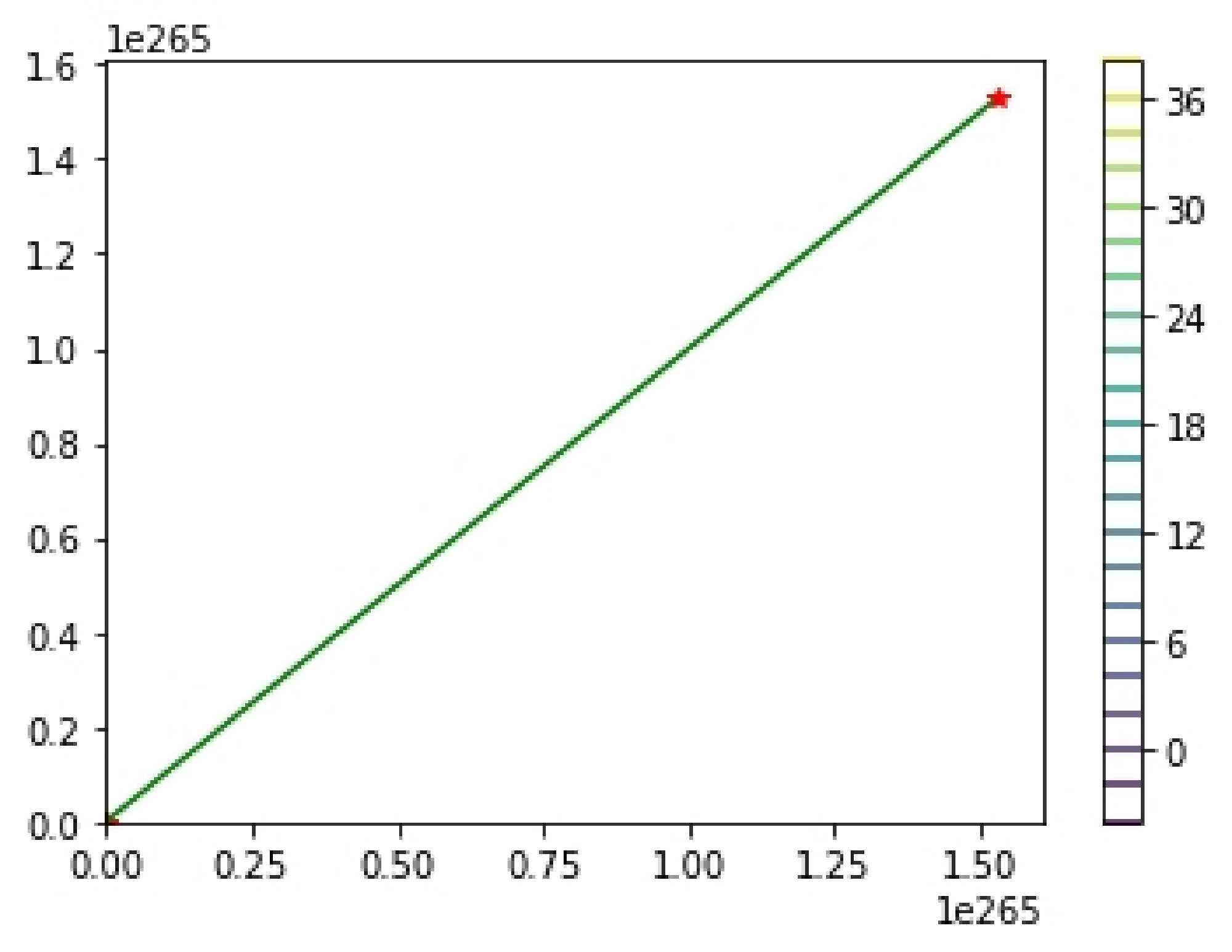
2 (a) $f(x) = \frac{1}{2}(Ax-b)^T(Ax-b) + \frac{\lambda}{2}\|x\|^2$
$\nabla f(x) = A^T(Ax-b) + \lambda x$
$\nabla^2 f(x) = A^T A + \lambda I$.

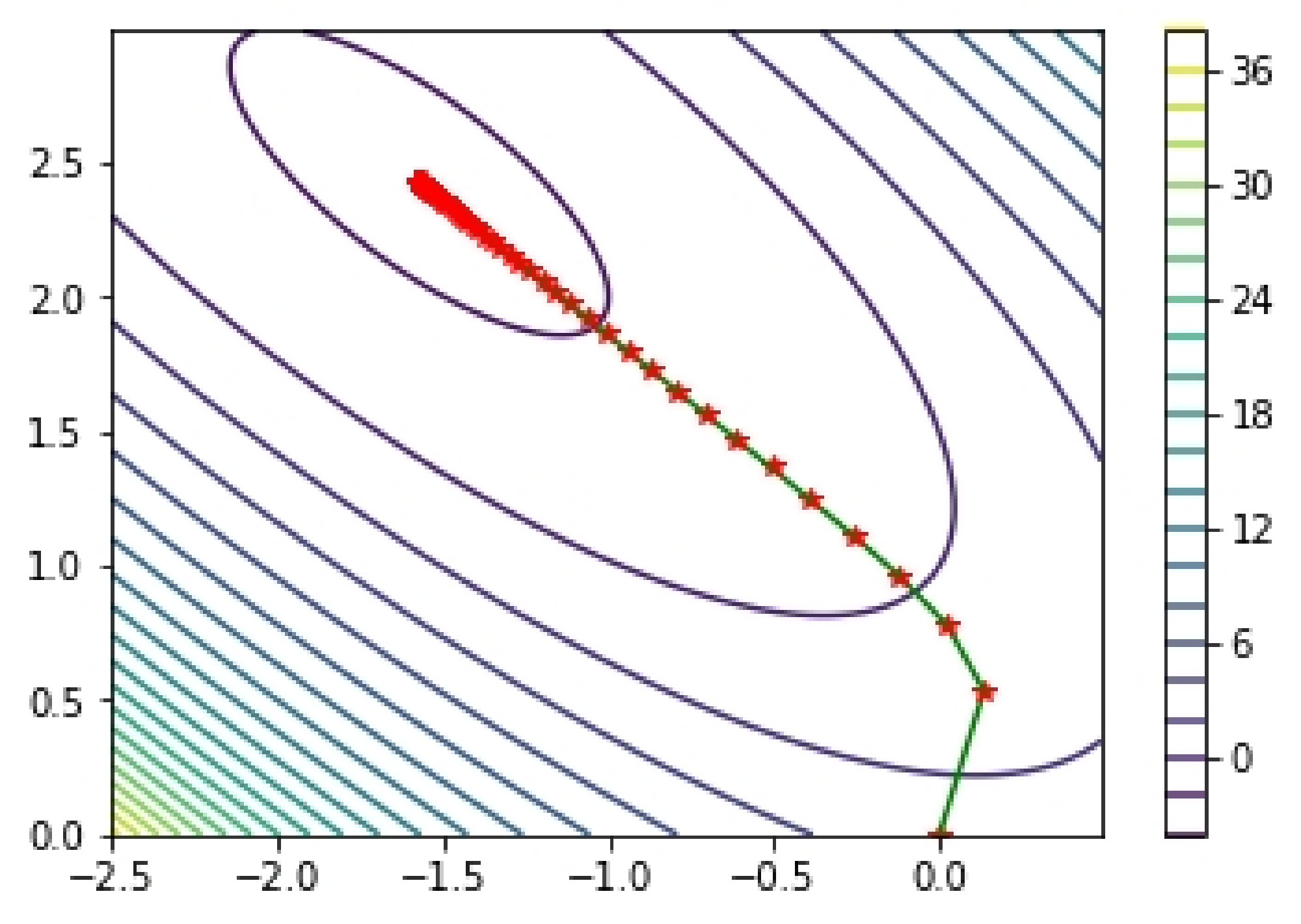
(b) $\nabla^2$ Hessian determines convexity.
Strongly convext: $A^T A$ is not positive definite $\lambda > 0$ ensure strong convexity
Convexity → $\nabla^2 f(x)$ positive semidefinite $\lambda \ge 0$
Non-convexity → $\nabla^2 f(x)$ not positive semidefinite $\lambda < \min(A^T A, 0)$

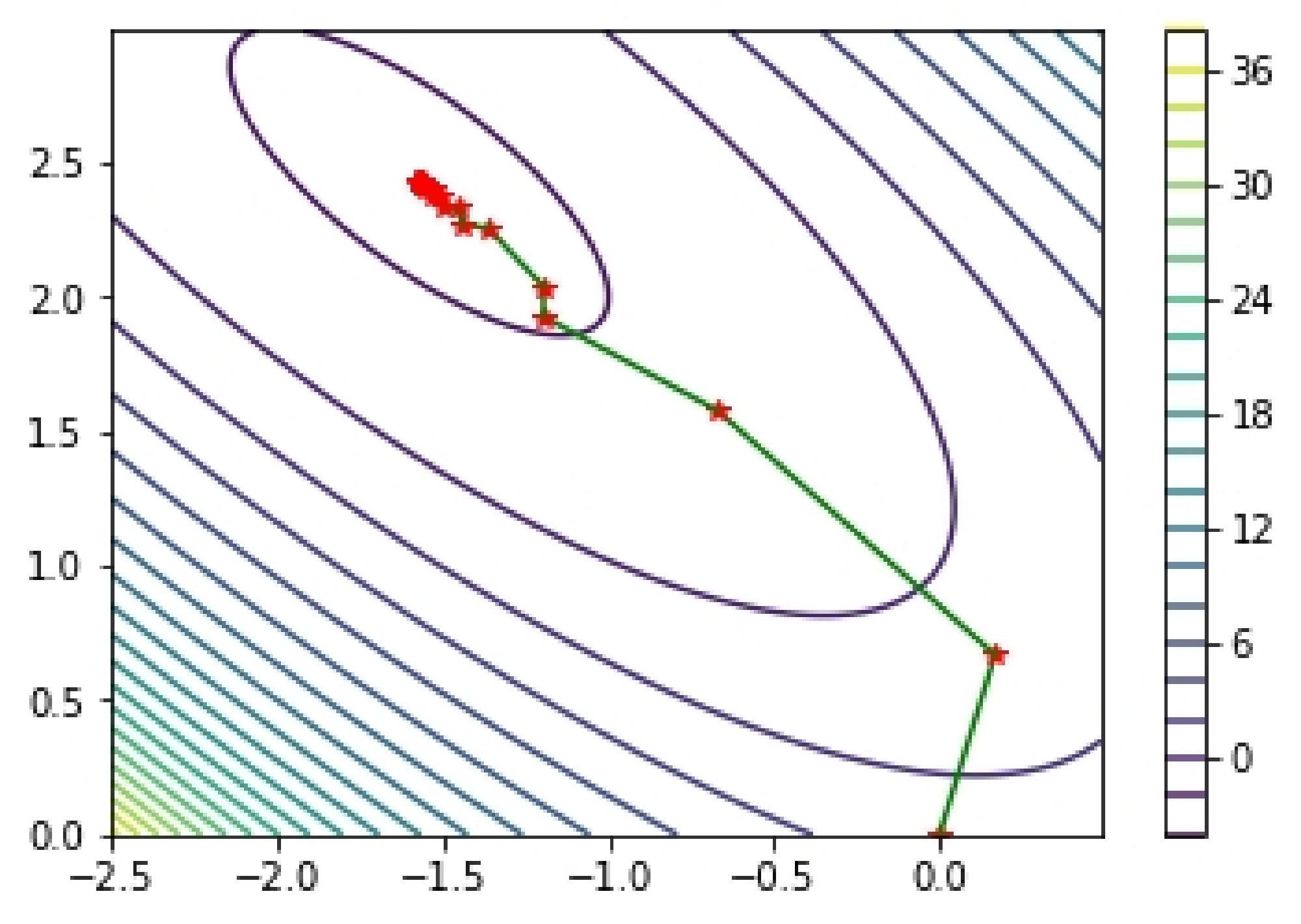
(c) ① $d^0 = -\nabla f(x^0) = -A^T(Ax-b) - \lambda x_{11}^0 = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$
$\varphi(\alpha) = f(x^0 + \alpha d^0) - f(x^0)$
$\varphi(\alpha) \le \gamma \alpha \varphi'(0)$, $x^0 + d^0 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} + \begin{pmatrix} -3 \\ -3 \end{pmatrix} = \begin{bmatrix} -\frac{5}{2} \\ -\frac{5}{2} \end{bmatrix}$
$f(x^0) = 1$ $f(x_i + \alpha d^0) = 91$ $\nabla f(x^0)^T d^0 = -9$ → not satisfied → continue.
Start with $\alpha = 1$ and we end at $\alpha' = 0.0625$
$x^1 = x^0 + \alpha'(d^0) = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + 0.0625 \begin{bmatrix} -3 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{13}{16} \\ -\frac{5}{16} \end{bmatrix}$.

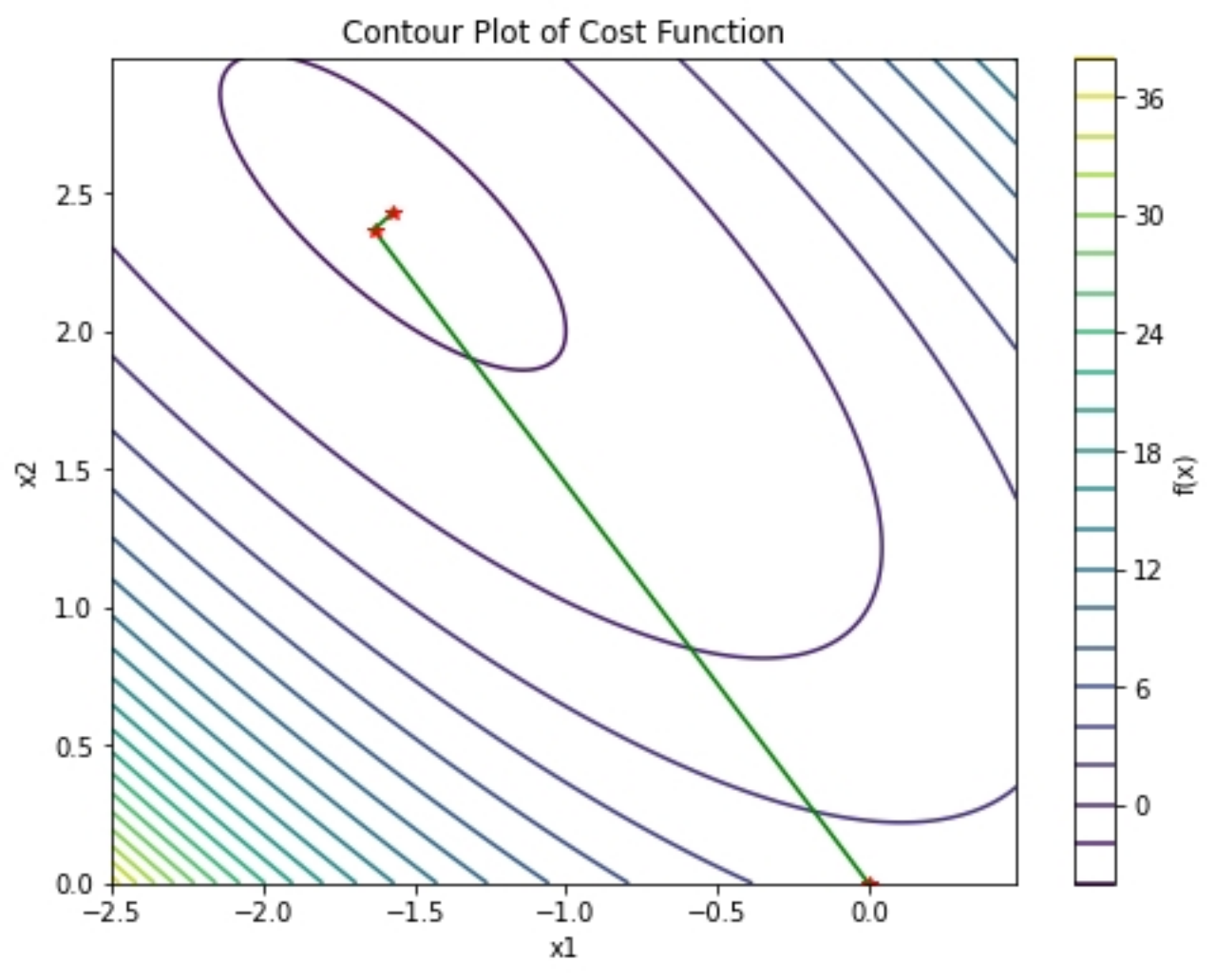
② $x^1 = x^0 - \alpha d^0 = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \frac{1}{12}\begin{bmatrix} -3 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{7}{12} \\ \frac{1}{12} \end{bmatrix}$
$d^1 = \nabla f(x^1)^T(-\nabla f(x^1)) = -\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \frac{7}{12} \\ \frac{1}{12} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \frac{3}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ not a descent direction

3. ① $\alpha = 1$ 5 iterations $x^* = [\text{nan}, \text{nan}]$ $f(x^*) = \text{nan}$ grad. norm = nan
② $\alpha = 0.1$ 120 $[-1.571284, 2.428703]$ $-2.28568$ $9.133687e-06$.
③ armijo. 41 $[-1.571286, 2.428704]$ $-2.285680$ $8.269390e-06$.
4. Newton 3 $[-1.571290, 2.428710]$ $-2.28568$ $2.167842e-07$


1e265
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1e265
36
30
24
18
12
6
0


2.5
2.0
1.5
1.0
0.5
0.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
36
30
24
18
12
6
0


2.5
2.0
1.5
1.0
0.5
0.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
36
30
24
18
12
6
0


Contour Plot of Cost Function
x1
x2
f(x)
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
0
6
12
18
24
30
36