1

ECE 236B - HW 7

Peter Raciappo
1039536 89

$\boxed{\text{A 4.26}}$ $\min \|Ax-b\|_2 + \gamma\|x\|_1$,

$A \in R^{m\times n}$, $b \in R^m$, $\gamma > 0$, $x \in R^n$.

(a.) $\min \|y\|_2 + \gamma\|x\|_1$
s.t. $Ax - b = y$

$$\mathcal{L}(x,y,\nu) = \|y\|_2 + \gamma\|x\|_1 + \nu^T(Ax-b-y)$$

$$g(\nu) = \inf_{x,y} \mathcal{L}(x,y,\nu) = \inf_{x,y} \|y\|_2 - \nu^T y + \gamma\|x\|_1 + \nu^T Ax - \nu^T b$$

$$\inf_y \|y\|_2 - \nu^T y = \begin{cases} 0, & \|\nu\|_2 \le 1 \\ -\infty, & \|\nu\|_2 > 1 \end{cases}$$

$$\inf_x \gamma(\|x\|_1 + \tfrac{1}{\gamma}\nu^T Ax) = \begin{cases} 0, & \|\tfrac{1}{\gamma}\nu^T A\|_\infty \le 1 \\ -\infty & \text{otherwise.} \end{cases}$$

$\|\frac{1}{\gamma}\nu^T A\|_\infty \le 1 \Rightarrow \|A^T\nu\|_\infty \le \gamma$
$\Rightarrow \max(|(A^T\nu)_i|) \le \gamma$, which is satisfied
if $|a_i^T\nu| \le \gamma$, $\forall i$, where a_i is the ith column of A.
($a_i = Ae_i$).

Thus, $g(\nu) = \begin{cases} -\nu^T b & \text{if } \|\nu\|_2 \le 1 \text{ and } |a_i^T\nu| \le \gamma, \forall i \\ -\infty & \text{o.w.} \end{cases}$

$|a_i^T\nu| \le \gamma \Leftrightarrow a_i^T\nu \le \gamma$ & $-a_i^T\nu \le \gamma$

The dual prob. is: $\max_\nu -\nu^T b$
s.t. $\|\nu\|_2 \le 1$ & $|a_i^T\nu| \le \gamma, \forall i$.

$\Rightarrow$ $\boxed{\begin{aligned} &\min\ b^T\nu \\ &\text{s.t. } \|\nu\|_2 \le 1 \\ &\qquad \|A^T\nu\|_\infty \le \gamma \end{aligned}}$

②

(b.) $Ax^* - b \neq 0$, x^* optimal.
Let $r = (Ax^* - b) / \|Ax^* - b\|_2$.

Note that $\|r\|_2 = 1$.

$Ax^* - b = y$ since x^* is feasible, $\Rightarrow r = y / \|y\|_2$.

KKT Conditions: $\nabla f_0 + \sum \nu \nabla h = 0$

$\frac{\partial}{\partial x_i} f_0 = \frac{\partial}{\partial x_i} \sum_i y_i^2 + \gamma |x_i| = \begin{cases} \gamma & \text{if } x_i > 0 \\ -\gamma & \text{o.w.} \end{cases}$

$\frac{\partial}{\partial y_i}(\,) = 2y_i$ $\quad \frac{\partial}{\partial x} h_i = \sum_j \frac{\partial}{\partial x_i} A_{ij} x_i = \sum_j A_{ij}$, $\quad \frac{\partial}{\partial y} h_i = -1$

$\Rightarrow 2y - \nu = 0 \quad \Rightarrow y = \frac{1}{2}\nu. \quad \Rightarrow \|y\|_2 = \frac{1}{2}\|\nu\|$

$\Rightarrow r = \nu / \|\nu\|_2$.

$\Rightarrow \|A^T r\|_\infty = \|A^T \nu\|_\infty / \|\nu\|_2 \leq \gamma / \|\nu\|_2 = \gamma$

KKT $\Rightarrow \gamma = A^T 1$

$r^T r = \frac{r^T(Ax^* - b)}{\|Ax^* - b\|} = \frac{r^T A x^* - b^T r}{\|Ax^* - b\|} = 1$

③

$\boxed{\text{A 4.28}}$ $\min \sum_{i=1}^{m} \sup_{a_i \in P_i} (a_i^T x - b_i)^2$.

$x \in \mathbb{R}^n$. $P_i = \{a_i \in \mathbb{R}^n \mid C_i a_i \leq d_i\}$,

$C_i \in \mathbb{R}^{p_i \times n}$, $d_i \in \mathbb{R}^{p_i}$.

$$\min \sum_{i=1}^{m} t_i^2$$
$$\text{s.t.} \quad \sup_{a_i \in P_i} \max\{a_i^T x - b_i, -a_i^T x + b_i\} \leq t_i, \quad i=1,\dots,m$$

$$g(\lambda_i) = \inf_{t_i, x} \mathcal{L}(t_i, x, \lambda_i) = \inf_{t_i, x} \left[t_i^2 + \lambda_i (\sup\max\{\ \} - t_i) \right]$$

$$= \inf_{t_i, x} \left[t_i^2 - \lambda_i t_i + \lambda_i \sup\max\{a_i^T x - b_i, -a_i^T x + b_i\} \right]$$

$$= \inf_{t_i} \left[t_i (t_i - \lambda_i) \right] + \lambda_i \inf_{x} \left[\sup_{a_i \in P_i} \max\{a_i^T x - b_i, -a_i^T x + b_i\} \right]$$

$$\frac{\partial}{\partial t_i} (t_i^2 - \lambda_i t_i) = 2t_i - \lambda_i = 0 \Rightarrow t_i = \tfrac{1}{2} \lambda_i$$

$$\Rightarrow \inf_{t_i} (\ ) = \min(0, \lambda_i / 2).$$

$$\sup_{a_i \in P_i} \max\{a_i^T x - b_i, -a_i^T x + b_i\} \leq t_i$$

$$\Updownarrow$$

$$\max\{a_i^T x - b_i, -a_i^T x + b_i\} \leq t_i, \quad \forall i$$
and $\quad C_i a_i \leq d_i$

$$\inf \max\{\ \} = 0$$

$\boxed{A\ 4.17}$ Let $f(A) = \sum_{k=1}^{r} \lambda_k(A)$.

(a.)

$$\begin{aligned} \max \quad & tr(AX) \\ s.t. \quad & tr X = r \\ & 0 \preceq X \preceq I, \quad X \in S^n. \end{aligned}$$

let $A = U^T \Lambda U$, $U^T U = I$, Λ diagonal.

$tr(AX) = tr(U^T \Lambda U X) = tr(\Lambda U X U^T) = tr(\Lambda Y)$,

where $Y = U X U^T$. Note: $tr(X) = tr(Y)$

Finally, $0 \preceq X \preceq I \Rightarrow 0 \preceq Y \preceq I$. So we have:

$$\begin{aligned} \max \quad & tr(\Lambda Y) \\ s.t. \quad & tr(Y) = r \\ & 0 \preceq Y \preceq I. \end{aligned}$$

$tr(\Lambda Y) = \sum_i (\Lambda Y)_{ii} = \sum_i (\Lambda_{ii} Y_{ii})$, since Λ diagonal.

$\Rightarrow tr(\Lambda Y) = tr(\Lambda \gamma)$, where $Y = Q^T \gamma Q$, $Q^T Q = I$.

$0 \preceq \gamma \preceq I \Rightarrow 0 \preceq Q^T \gamma Q \preceq Q^T Q \Rightarrow 0 \preceq Y \preceq I$.

$0 \preceq Y \preceq I \Rightarrow 0 \preceq Q Y Q^T \preceq Q Q^T \Rightarrow 0 \preceq \gamma \preceq I$.

Thus, $0 \preceq Y \preceq I \Leftrightarrow 0 \preceq \gamma \preceq I$. We have:

$$\begin{aligned} \max \quad & tr(\Lambda \gamma) \\ s.t. \quad & tr(\gamma) = r \\ & 0 \preceq \gamma \preceq I, \\ & \Lambda, \gamma \text{ diagonal}. \end{aligned}$$

$$\Rightarrow \quad \begin{aligned} \max \quad & \textstyle\sum_i \Lambda_{ii} \gamma_{ii} \\ s.t. \quad & \textstyle\sum_i \gamma_{ii} = r \\ & 0 \le \gamma_{ii} \le 1, \ \forall i \end{aligned}$$

The optimal solution of this problem is clearly to set $\gamma_{ii} = 1$ for all $i \in [1, r]$, since $\Lambda_{ii} > 0$ for any i. Thus, the optimal solution is $\sum_i^r \Lambda_{ii} = \sum_{k=1}^{r} \lambda_k(A) = f(A)$.

(5)

(b.) $f(A) = \sum_{k=1}^{r} \lambda_k(A) = \sum_{k=1}^{r} \lambda_k(\Lambda)$

$= \sum_{k=1}^{r} \lambda_{max}(\Lambda_k)$, where $\Lambda_k = \begin{bmatrix} \lambda_k & & 0 \\ 0 & 0 & \\ & & \ddots \; 0 \end{bmatrix}$.

$= \sum_{k=1}^{r} \sup_{\|y\|_2=1} y^T \Lambda_k y$, which is convex.

(c.) Assume $A(x) = A_0 + x_1 A_1 + \quad + x_m A_m$, $A_k \in S^n$

$\min f(A(x)) = \min \sum_{k=1}^{r} \lambda_k(A_0 + x_1 A_1 + \cdots + x_m A_m)$

$= \min \text{tr}(A_0' + x_1 A_1' + \cdots + x_m A_m')$. ?

Dual of (a.):
$$\begin{aligned} &\min -\text{tr}(AX) \\ &\text{s.t. } \text{tr}\, X = r \\ &\quad 0 \preceq X \preceq I. \end{aligned}$$

$g(\nu, Z, P) = \inf_X [-\text{tr}(AX) + \nu(\text{tr}X - r) + \text{tr}((X-I)Z) + \text{tr}(-XP)]$

$= \inf_X [-\text{tr}(AX) + \text{tr}(\nu X) + \text{tr}(XZ) - \text{tr}(Z) - \text{tr}(XP) - \nu r]$

$= \inf_X [\text{tr}((-A + I\nu + Z - P)X)] - \text{tr}(Z) - \nu r$

$= \begin{cases} -\text{tr}(Z) - \nu r & \text{if } -A + I\nu + Z - P = 0 \\ -\infty & \text{o.w.} \end{cases}$

$\Rightarrow$
$$\begin{aligned} &\min \text{tr}(Z) + \nu r \\ &\text{s.t. } -A + I\nu + Z - P = 0 \\ &\quad Z \succeq 0,\ P \succeq 0 \end{aligned}$$

$\Rightarrow$
$$\begin{aligned} &\min \text{tr}(Z) + \nu r \\ &\text{s.t. } 0 \preceq Z \preceq A - I\nu \end{aligned}$$

⑥

So we have:

$$\boxed{\begin{aligned} &\min \ \mathrm{tr}(Z) + \nu r \\ &\text{s.t.} \ 0 \preceq Z \preceq A_0 - I\nu + x_1 A_1 + \dots + x_n A_n \end{aligned}},$$

which is an SDP. By convexity of the problem, strong duality holds, so the solution minimizes $f(A(x))$.

$\boxed{4.10}$ $\min \|Ax - b\|_2^2$

s.t. $x_k^2 = 1$, $k \in [1, n]$.

$\mathrm{rank}(A) = n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$.

(a.) <u>Dual</u>

$$\mathcal{L}(x, \nu) = \|Ax - b\|_2^2 + x^T \mathrm{diag}(\nu) x - \nu^T 1$$

$$g(\nu) = \inf_x (Ax-b)^T(Ax-b) + x^T \mathrm{diag}(\nu) x - \nu^T 1$$

$$= \inf_x \left(x^T A^T A x - 2b^T A x + b^T b + x^T \mathrm{diag}(\nu) x - \nu^T 1 \right)$$

$$= \inf_x \left(x^T (A^T A + \mathrm{diag}(\nu) x - 2b^T A x \right) + b^T b - \nu^T 1.$$

$$= \begin{cases} (\ ?\ ) & \text{if } A^T A + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{o.w.} \end{cases}$$

More precisely, $\succeq 0$ iff $\|(A^TA + \mathrm{diag}(\nu))A^T b\|$

$$\frac{\partial}{\partial x}\left(x^T (A^T A + \mathrm{diag}(\nu)) x - 2b^T A x \right)$$

$$= 2(A^T A + \mathrm{diag}(\nu)) x - 2A^T b = 0$$

$\Rightarrow x = (A^T A + \mathrm{diag}(\nu))^{-1} A^T b$. Note: $(A^T A + \mathrm{diag}(\nu))$ is symmetric

$$\Rightarrow g(\nu) = b^T A (A^T A + \mathrm{diag}(\nu))^{-T} (A^T A + \mathrm{diag}(\nu))(A^T A + \mathrm{diag}(\nu))^{-1} A^T b$$

$$- 2b^T A (A^T A + \mathrm{diag}(\nu))^{-1} A^T b + b^T b - \nu^T 1.$$

$$= -b^T A (A^T A + \mathrm{diag}(\nu))^{-1} A^T b + b^T b - \nu^T 1.$$

7

In summary,

$$g(\nu) = \begin{cases} -b^TA(A^TA + \mathrm{diag}(\nu))^{-1}A^Tb + b^Tb - \nu^T\mathbf{1}, \\ \qquad \text{if } A^TA + \mathrm{diag}(\nu) \succeq 0. \\ -\infty \quad \text{otherwise} \end{cases}$$

The dual is:

$$\max \quad b^Tb - \nu^T\mathbf{1} - t$$
$$\text{s.t.} \quad t \geq b^TA(A^TA + \mathrm{diag}(\nu))^{-1}A^Tb$$
$$\Rightarrow \begin{bmatrix} A^TA + \mathrm{diag}(\nu) & A^Tb \\ b^TA & t \end{bmatrix} \succeq 0$$

So we have,

$$\boxed{\begin{array}{ll} \max\limits_{\nu,t} & b^Tb - \mathbf{1}^T\nu - t \\ \text{s.t.} & \begin{bmatrix} A^TA + \mathrm{diag}(\nu) & A^Tb \\ b^TA & t \end{bmatrix} \succeq 0 \\ & \begin{cases} A^TA + \mathrm{diag}(\nu) \succeq 0 \\ \nu \succeq 0, \ t \geq 0 \end{cases} \end{array}}$$

Implicit constraints

(b.) Expressing this as a minimization:

$$\min \quad \mathbf{1}^T\nu + t - b^Tb$$
$$\text{s.t.} \quad -\begin{bmatrix} A^TA + \mathrm{diag}(\nu) & -A^Tb \\ -b^TA & -t \end{bmatrix} \preceq 0$$

(8)

$$\mathcal{L}(\nu, t, Z, q, p) = 1^T\nu + t - b^Tb - tr\left(\begin{bmatrix} Z & q \\ q^T & p \end{bmatrix}\begin{bmatrix} A^TA + diag(\nu) & -A^Tb \\ -b^TA & t \end{bmatrix}\right)$$

$$= 1^T\nu + t - b^Tb - tr\left(\begin{bmatrix} Z(A^TA + diag(\nu)) - qb^TA & \times \\ \times & -q^TA^Tb + pt \end{bmatrix}\right)$$

$$= 1^T\nu + t - b^Tb - \left[tr(Z(A^TA + diag(\nu))) - 2q^TA^Tb + pt\right]$$

$$= 1^T\nu + t - b^Tb - tr\left[Z(A^TA + diag(\nu))\right] + 2q^TA^Tb - pt$$

$$= 1^T\nu - tr(Z\,diag(\nu)) + t - b^Tb - tr(ZA^TA) + 2q^TA^Tb - pt.$$

$$= (1 - diag(Z))^T\nu + (1-p)t - b^Tb - tr(ZA^TA) + 2q^TA^Tb$$

$$g(t, Z, q, p) = \inf_{\nu, t} \; (1 - diag(Z))^T\nu + (1-p)t + (\ldots)$$

$$= \begin{cases} -b^Tb - tr(ZA^TA) + 2q^TA^Tb & \text{if } diag(Z) = 1,\ p = 1. \\ -\infty & \text{otherwise.} \end{cases}$$

"p" doesn't appear and can be eliminated.

Thus, the dual is:

$$\begin{aligned} \max \;& -b^Tb - tr(ZA^TA) + 2q^TA^Tb \\ \text{s.t.} \;& diag(Z) = 1 \end{aligned}$$

$\Leftrightarrow$

$$\begin{aligned} \min \;& tr(A^TAZ) - 2b^TAz + b^Tb \\ & diag(Z) = 1 \end{aligned}$$

$\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0 \Leftrightarrow Z \succeq 0$ and $z \succeq 0$. So we have:

$$\begin{aligned} \min \;& tr(A^TAZ) - 2b^TAz + b^Tb \\ & diag(Z) = 1 \\ & \begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0 \end{aligned}$$

9

Rewrite $\min \|Ax-b\|_2^2$

$\text{s.t. } x_k^2 = 1, \; k \in [1,n]$.

$\min \; x^T A^T A x - 2b^T A x + b^T b \quad \text{s.t. } \operatorname{diag}(xx^T) = 1$.

Let $z = x$ and let $Z = zz^T = xx^T$.

$\Rightarrow x^T A^T A x = tr(x^T A^T A x)$ (since $x^T A^T A x$ is a scalar)

$= tr(A^T A x x^T) = tr(A^T A Z)$, so we have:

$$\min \; tr(A^T A Z) - 2b^T A z + b^T b$$
$$\text{s.t. } \operatorname{diag}(Z) = 1$$
$$Z = zz^T$$

In the dual,

we have $\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0 \iff \begin{cases} z^T Z^{-1} z \leq 1, \\ zz^T \preceq Z \end{cases}$

So we have relaxed the constraint $Z = zz^T$ to $Z \succeq zz^T$.

Suppose that $\operatorname{rank}\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} = 1$.

Then $\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} = \begin{bmatrix} v \\ t \end{bmatrix} [v^T \; t]$ for some v, t.

$\Rightarrow \begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} = \begin{bmatrix} vv^T & tv \\ tv^T & t^2 \end{bmatrix} \qquad t^2 = 1 \Rightarrow t = 1 \text{ or } -1.$

$\Rightarrow z = v \text{ or } z = -v$

In either case $Z = vv^T \Rightarrow Z = zz^T$,

So the problems are the same and have the same optimal value.

10

(c.) Let $v \in \mathbb{R}^n$, $z = Ev$, $Z = Evv^T$.

Consider $\min E\|Av - b\|_2^2$
s.t. $E v_k^2 = 1$.

$\Rightarrow \min E(v^TA^TAv - 2b^TAv + b^Tb)$
s.t. $E \operatorname{diag}(vv^T) = 1$.

$E(v^TA^TAv) = E\sum_i \sum_j A_{ji}A_{ij} v_i^2 = \sum_{i,j} A_{ij}A_{ji} E(v_i^2)$

$= \sum_{ij} A_{ij}A_{ji} Z_{ii} = A^TAZ$. More simply,

$E(v^TA^TAv) = E(tr(v^TA^TAv)) = E(tr(vv^TA^TA))$
$= tr(E(vv^T)A^TA) = tr(ZA^TA)$.

$\Rightarrow E\|Av-b\|_2^2 = tr(ZA^TA) - 2b^TAz + b^Tb$.

Also, $E(\operatorname{diag}(vv^T)) = \operatorname{diag}(E(vv^T)) = \operatorname{diag}(Z)$.

$Z \succeq 0$, $z \geq 0$ $\Leftrightarrow$ $\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0$. So we again have:

$$\begin{aligned} \min\ & tr(ZA^TA) - 2b^TAz + b^Tb \\ \text{s.t. } & \operatorname{diag}(Z) = 1 \\ & \begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0. \end{aligned}$$

In [1]:
```
import numpy as np
import cvxpy as cp
import matplotlib.pyplot as plt
```

4.10

In [93]:
```
m = 50
n = 40

for s in np.array((0.5,1,2,3)):
    A = np.random.uniform(0,1,(m,n))
    xhat = np.sign(np.random.uniform(0,1,n))
    b = np.matmul(A,xhat) + s*np.random.uniform(0,1,m)
#     np.linalg.matrix_rank(A)

    z = cp.Variable(n)
    Z = cp.Variable((n,n), symmetric=True)
    # Q = cp.bmat([[Z,z],[z.T,1]])
    Q = np.array([[Z,z],[z.T,1]])[0][0]
    # u = np.hstack((Z,z))
    # l = np.hstack((z.T,1))
    # Q = np.vstack((u,l))[0][0]

    # constraints = [cp.diag(Z) == 1]
    # constraints = [cp.diag(Z) == np.ones(n)]
    # constraints += [Z >> 0]
    # constraints += [z >= 0]
    # constraints += [Z >> z @ z.T]
    # constraints += [cp.bmat([[Z,z],[z.T,1]]) >> 0]

    constraints += [np.array([[Z,z],[z.T,1]]) >> 0]
    constraints += [Q >> 0]
    objective = cp.trace(Z @ A.T @ A) - 2*b.T @ A @ z + b.T @ b
    prob = cp.Problem(cp.Minimize(objective),constraints)
    prob.solve()
    print("Exact results:")
    print(np.round(prob.value,2))

    # Define and solve the CVXPY problem.
    x = cp.Variable(n)
    cost = cp.sum_squares(A @ x - b)
    prob = cp.Problem(cp.Minimize(cost))
    prob.solve()

    print("(i)")
    print(np.sqrt(np.sum((np.matmul(A,np.sign(x.value)) - b)**2)))

    print("(ii)")
```

```
    print(np.round(cp.trace(Z.value @ A.T @ A) - 2*b.T @ A @ z.value + b.T @ b),2)
```

```
Exact results:
4.05
7.87
15.2
22.11
(i)
2.083638602553853
4.195202877937251
34.16791929347048
54.968165926921806
(ii)
4.16
8.32
16.65
25.96
```

5.4

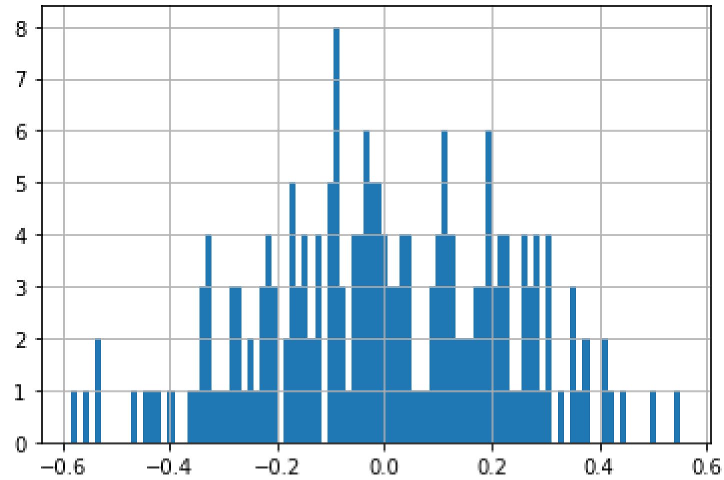
In [99]:

```
m = 200
n = 100
A = np.random.uniform(0,1,(m,n))
b = np.random.uniform(0,1,m)
b = b/(1.01*max(abs(b)))
```

(a)

In [135]:
```
x = np.dot(np.linalg.pinv(A),b)
x.T

plt.hist(np.matmul(A,x)-b,bins=int(m/2))
plt.grid()
plt.show()
```

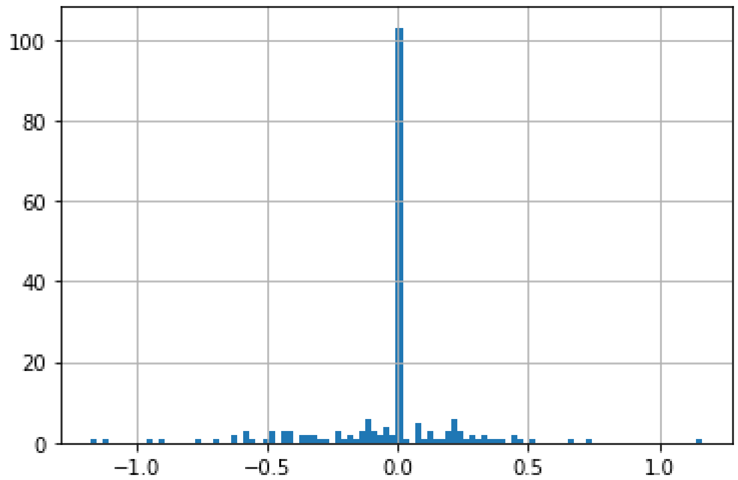


(b)

In [176]:
```
x = cp.Variable(n)
cost = cp.norm(A @ x - b,1)
prob = cp.Problem(cp.Minimize(cost))
prob.solve()
# print(x.value)
print(prob.value)

plt.hist(np.matmul(A,x.value)-b,bins=int(m/2))
plt.grid()
plt.show()
```

```
30.52012086553837
```

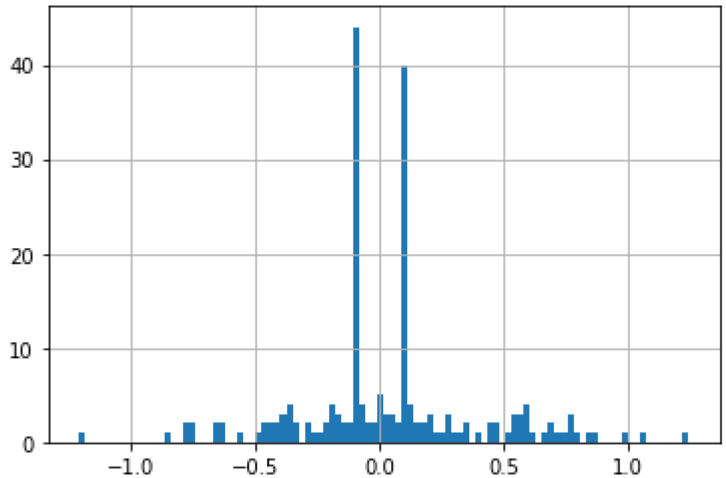


(d)

In [174]:
```
x = cp.Variable(n)

term1 = np.zeros(m)
term2 = cp.abs(A @ x - b) - 0.2
term3 = 2*cp.abs(A @ x - b) - 0.5

cost = cp.sum(cp.maximum(term1,term2,term3))
# cost = 0
# for i in np.arange(m):
#     cost += cp.sum(cp.maximum(term1[i],term2[i],term3[i]))
prob = cp.Problem(cp.Minimize(cost))
prob.solve()
# print(x.value)
print(prob.value)
```



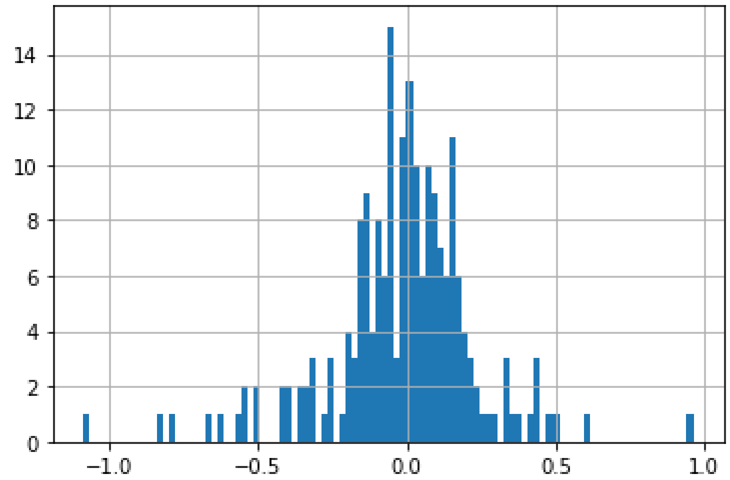
(e)

In [175]:
```
x = cp.Variable(n)

cost = cp.sum(cp.huber(A @ x - b, 0.2))
prob = cp.Problem(cp.Minimize(cost))
prob.solve()
# print(x.value)
print(prob.value)

plt.hist(np.matmul(A,x.value)-b,bins=int(m/2))
plt.grid()
plt.show()
```

7.803564375786363



In [ ]: