1(a) \[ \left\| x - p(x) - y - p(y) \right\|^2 \]
\[ = \left\| x - y \right\|^2 - 2 \left\langle x - y, (p(x) - p(y)) \right\rangle + \left\| p(x) - p(y) \right\|^2 \]
\[ = \left\| x - y \right\|^2 - 2 \left( x - y \right)^T (p(x) - p(y)) + \left\| p(x) - p(y) \right\|^2 \]
\[ = \left\| x - y \right\|^2 - 2 \frac{(x - p(x))^T (p(x) - p(y))}{\left\| p(x) - p(y) \right\|^2} \times 0 \]
\[ = \left\| x - y \right\|^2 - 2 \frac{(y - p(y))^T (p(x) - p(y))}{\left\| p(x) - p(y) \right\|^2} \times 0 \]
\[ \leq \left\| x - y \right\|^2 - \left\| p(x) - p(y) \right\|^2 \]
\[ \leq \left\| x - y \right\|^2 \]

1(b) Define \( \overline{f} = -p(x) \) and let \( \alpha = \left\| \overline{f} \right\|^2 \).

For all \( x \in \mathbb{R}^2 \) have:
\[ (x - p(x))^T (p(x) - p(0)) \leq 0 \]
\[ \Rightarrow (x + \overline{f})^T \overline{f} \leq 0 \]
\[ \Rightarrow \overline{f}^T \overline{f} = -\alpha \]

2(a) Suppose \[ \begin{bmatrix} \nabla^2 f(x^*) - \nabla^2 g(y) \\ \nabla g(y)^T \\ 0 \end{bmatrix} u = 0 \]

We have \( \nabla g(y)^T u = 0 \).

From first row, \( \nabla^2 f(x^*) \cdot u = \nabla^2 g(y) \cdot u = 0 \).
\[ \Rightarrow u^T \nabla^2 f(x^*) \cdot u = 0 \]
\[ \Rightarrow u^T \nabla^2 f(x^*) \cdot u = 0 \]
\[ \Rightarrow u^T \nabla^2 f(x^*) \cdot u \geq 0 \]
\[ \Rightarrow u = 0, \text{ since } u \in \mathbb{R}^n \]

Thus from first row \( u = 0 \).

Since \( \nabla f(x^*) \) has full column rank, we have \( u = 0 \). Hence \( (u, v) = 0 \), so the matrix is nonsingular.
Let $u$ be a solution with some initial set $A$ as

for original problem. It must satisfy

$$\nabla E(x) = \nabla (\mathcal{A}(t)) x + \varepsilon E$$

$\Rightarrow \mathcal{A}(x) = \varepsilon \mathcal{E}$

We can apply the implicit function theorem to

the following system of parameterized nonlinear equations

$$h(x, x_0, \varepsilon) = \begin{bmatrix} \nabla f(x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b \end{bmatrix} = 0$$

Because $h(x, x_0, \varepsilon) = 0$, $\frac{\partial h}{\partial \varepsilon} = 0$

is nowhere zero at $(x, x_0) = (x, x_0)$, and $f = 0$ (in part (b))

Also, we have.

$$h(x^2, x^2, 0) = 0$$

Applying the implicit function theorem, we conclude

that $(x^2, x^2, 0)$ is a smooth function of $\varepsilon$.

Thus

$$\begin{align*}
\frac{d}{d\varepsilon} \left[ \begin{array}{c} x^2 \\ x^2
\end{array} \right] &= -\begin{bmatrix} \nabla f(x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}^{-1} \begin{bmatrix} \nabla f(x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}
\\ &= -\begin{bmatrix} \nabla^2 f(x, x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b \\
\nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}^{-1} \begin{bmatrix} \nabla f(x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}
\\ &= -\begin{bmatrix} \nabla^2 f(x, x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b \\
\nabla \mathcal{A}(x) \mathcal{A}(x) + \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}^{-1} \begin{bmatrix} \nabla f(x) - \nabla \mathcal{A}(x) \mathcal{A}(x) + \varepsilon b
\end{bmatrix}
\end{align*}$$

Thus $x^2 = x^2 + \varepsilon \mathcal{A}(x) + \sigma(\varepsilon)$

where $g = \mathcal{A}(x) + \varepsilon \mathcal{A}(x)^T + \sigma(\varepsilon)$.
3. Let \( B = \sum_{i=0}^{n} \lambda_i q_i q_i^T \) be the eigenvalue position
\[
A \cdot B = \sum_{i=0}^{n} \lambda_i \cdot A \cdot (q_i q_i^T)
\]
\[
= \sum_{i=0}^{n} \lambda_i \cdot \text{trace}(A q_i q_i^T)
\]
\[
= \sum_{i=0}^{n} \lambda_i \cdot q_i^T A q_i
\]

Suppose \( A \cdot B \geq 0 \) for all \( A \geq 0 \); If we choose
\[
A = q_j q_j^T \quad \text{for some } j \quad \text{we obtain}
\]
\[
A \cdot B = \lambda_j (q_j q_j^T)^2 = \lambda_j \geq 0
\]
Hence \( \lambda_j > 0 \) \( \forall j \), so \( B \geq 0 \).

Conversely, if \( B \geq 0 \), we have \( \lambda_j > 0 \) \( \forall j \),
and additionally \( q_j^T A q_j \geq 0 \), \( j = 1, \ldots, n \), so
\[
A \cdot B = \sum_{i=0}^{n} \lambda_i \cdot q_i^T A q_i \geq 0
\]
Since the change in $\mu$ for a step of length $l$ is
\[ \mu(l) - \mu(l - \sigma) \],
we maximize the decrease in $\mu$ by minimizing $\sigma$. Thus, for a given $\theta \in (0, 1)$, we seek the smallest $\sigma$ that satisfies the condition
\[ \frac{\sigma^2 + n(1 - \sigma)^2}{2 \sqrt{2} - n} \leq \theta. \]

We rearrange to obtain a quadratic in $\sigma$.
\[ \sigma^2 + n(1 - 2\sigma + \sigma^2) \leq \frac{\sigma^2}{2} \frac{2 \sqrt{2}}{n} (1 - \theta) \]
\[ \sigma^2 - \sigma(2n + 2 \sqrt{2} \theta (1 - \theta)) + (\theta^2 + n) \leq 0 \]

The quadratic on the left is convex, and the inequality is not satisfied at $\sigma = 0$. The roots of the quadratic are
\[ \sigma = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \]
\[ = \frac{-2n \theta (1 - \theta) \pm \sqrt{(2n \theta (1 - \theta))^2 - 4n \theta^2 (\theta^2 + n)}}{2n} \]

Clearly, both roots, if they exist, are positive.
- If \( \left(1 + \frac{2 \theta (1 - \theta)}{n}\right)^2 < (1 + \theta^2) \), then no roots exist and there are no $\sigma$ satisfying the inequality.
- This may happen if $\theta$ is close to $0$ (but smaller than $1$).
- If both roots exist but both are greater than $1$, we again have no suitable choice for $\theta$.

This happens when
\[ \left(1 + \frac{2 \theta (1 - \theta)}{n}\right) - \left(1 + \frac{\sqrt{2} \theta (1 - \theta)}{n}\right) - \left(1 + \frac{\theta^2}{n}\right) \geq 1 \]
which after rearrangement and simplification is equivalent to

\[ 2 \frac{1}{\epsilon} (1 - \Theta) - 2^2 \leq 0 \]

Again, this may happen if \( \Theta \) is close to (but smaller than) 1.

- Otherwise, the smaller of the two roots is the "optimal" choice for \( \epsilon \).
2. (a) 
\( \begin{align*} 
(A) \quad & g + \mu x - A^T u = 0, \quad 0 \leq u \perp Ax - b \geq 0. \\
(B) \quad & g - A^T v + 2\gamma z = 0, \quad 0 \leq v \perp Az = b \geq 0, \quad 0 \leq \gamma \perp \Delta^2 - z^T z \geq 0. 
\end{align*} \)

(b) For (B), since \( b \leq 0 \), we note that the feasible set contains 0 and is thus nonempty. Moreover, it is closed and bounded, because of the condition \( ||z||_2 \leq \Delta \). Hence, the problem is one of minimizing a continuous (in fact, linear) function over a compact set, so it attains a solution.

For (A), we note that 0 is a feasible point, so that we can add the constraint \( g^T x + (\mu/2)x^T x \leq g^T 0 + (\mu/2)0^T 0 = 0 \) to obtain an equivalent formulation. This additional constraint defines a bounded set, since
\[
\frac{\mu}{2} ||x||^2 \leq -g^T x \leq ||g|| ||x|| \implies ||x|| \leq \frac{2}{\mu} ||g||. 
\]

Thus the modified problem reduces to minimization of a smooth (quadratic) function over a compact set, so again the minimum is attained.

(c) Let \( x(\mu) \) solve (A) with Lagrange multiplier \( u(\mu) \). If we set \( \Delta = ||x(\mu)||_2^2 \), then \( x(\mu) \) satisfies the KKT conditions for (B) with \( v = u(\mu) \) and \( \gamma = (1/2)\mu \).

(d) Choose \( \mu_1 \) and \( \mu_2 \) with \( 0 < \mu_1 < \mu_2 \). Since \( x(\mu_2) \) is not optimal for problem (A) when \( \mu = \mu_1 \), we have
\[
g^T x(\mu_1) + \frac{\mu_1}{2} x(\mu_1)^T x(\mu_1) \leq g^T x(\mu_2) + \frac{\mu_1}{2} x(\mu_2)^T x(\mu_2). 
\]

Similarly, since \( x(\mu_1) \) is not optimal for (A) when \( \mu = \mu_2 \), we have
\[
g^T x(\mu_2) + \frac{\mu_2}{2} x(\mu_2)^T x(\mu_2) \leq g^T x(\mu_1) + \frac{\mu_2}{2} x(\mu_1)^T x(\mu_1). 
\]

By adding these two inequalities and doing some elementary manipulation, we obtain
\[
\frac{\mu_2 - \mu_1}{2} x(\mu_1)^T x(\mu_1) \leq \frac{\mu_2 - \mu_1}{2} x(\mu_2)^T x(\mu_2). 
\]

Since \( \mu_2 > \mu_1 \), this formula implies that \( ||x(\mu_1)||_2 \leq ||x(\mu_2)||_2 \), proving the result.