① Let's look at:

$$(x(t+1)-x^*)^T \Lambda^{-1} (x(t+1)-x^*) =$$

$$(x(t+1)-x(t)+x(t)-x^*)^T \Lambda^{-1} (x(t+1)-x(t)+x(t)-x^*) =$$

$$= (x(t)-x^*)^T \Lambda^{-1} (x(t)-x^*) + 2(x(t+1)-x(t))^T \Lambda^{-1} (x(t)-x^*)$$

$$+ (x(t+1)-x(t))^T \Lambda^{-1} (x(t+1)-x(t))$$

By the update rule we know that:

$$(x(t+1)-x(t))^T \Lambda^{-1} (x(t)-x^*) = -(\gamma \Lambda \nabla f(x(t)))^T \Lambda^{-1} (x(t)-x^*) =$$

$$= -\gamma \nabla f(x(t))^T \underbrace{\Lambda^T \Lambda^{-1}}_{I} (x(t)-x^*) = -\gamma [\nabla f(x(t)) - \underbrace{\nabla f(x^*)}_{=0}](x(t)-x^*)$$

Assuming Λ is diagonal or just that $\Lambda^T = \Lambda$.

$$\leq -\frac{\gamma}{L} \|\nabla f(x(t)) - \nabla f(x^*)\|^2 = -\frac{\gamma}{L} \|\nabla f(x(t))\|$$

by the Lipschitz assumption

Also:

$$(x(t+1)-x(t))^T \Lambda^{-1} (x(t+1)-x(t)) = \gamma^2 [\nabla f(x(t))]^T \Lambda^T \nabla f(x(t))$$

Next, note that:

$(x(t+1)-x^*)^T \Lambda^{-1} (x(t+1)-x^*)$ and $(x(t)-x^*)^T \Lambda^{-1} (x(t)-x^*)$

are just scaled norms, so all I have to do is show that as $t \to \infty$, $(x(t)-x^*)^T \Lambda^{-1} (x(t)-x^*) \longrightarrow 0$ (then $x(t)$ converges to the minimizer x^*).

Rearranging the previous inequality I get:
(and using the inequalities I got)

$$(x(t)-x^*)^T\Lambda^{-1}(x(t)-x^*)-(x(t+1)-x^*)^T\Lambda^{-1}(x(t+1)-x^*) \geq$$

$$\geq 2\frac{\gamma}{L}\|\nabla f(x(t))\|^2 - \gamma^2 \nabla f(x(t))^T \underbrace{\Lambda^T}_{\substack{=\Lambda \\ \text{assume}}} \nabla f(x(t)) =$$

$$= \|\nabla f(x(t))\|^2 \left(2\frac{\gamma}{L} - \gamma^2 \underbrace{\frac{\nabla f(x(t))^T\Lambda^T\nabla f(x(t))}{\|\nabla f(x(t))\|^2}}\right) \geq$$

by the properties of Λ, we know $\forall x$, $\frac{x^T\Lambda x}{x^T x} \leq \lambda_{max}$ since Λ is positive definite

$$\geq \overbrace{\|\nabla f(x(t))\|}^{\geq 0} \underbrace{\left(2\frac{\gamma}{L} - \gamma^2\lambda_{max}\right)}_{>0 \text{ because}} > 0$$

by assumption

$$2\frac{\gamma}{L}-\gamma^2\lambda_{max} > 0 \Leftrightarrow 2\frac{\gamma}{L} > \gamma^2\lambda_{max} \Leftrightarrow \frac{2}{L\lambda_{max}} > \gamma$$ true by assumption

So this shows that

$(x(t)-x^*)^T\Lambda^{-1}(x(t)-x^*) > (x(t+1)-x^*)\Lambda^{-1}(x(t+1)-x^*)$, so as $t\to\infty$, this is strictly decreasing to 0, so $x(t)-x^*\to 0$ so $x(t)\to x^*$, so this shows $x(t)$ converges to x^*, i.e. the minimizer of $f(\cdot)$ $\square$.

(2) The objective function is:

$$f(w) = \|Aw - y\|^2 + \lambda \|w\|^2$$, where $\|\cdot\|$ is the L-2 norm (for the context of this problem).

Then:

a) To show f is smooth, I WTS $\forall u, v$ $\|\nabla f(u) - \nabla f(v)\| \leq L \|u - v\|$

Then:

$$\nabla_w f(w) = \nabla f(w) = \nabla\left(\|Aw - y\|^2 + \lambda \|w\|^2\right) =$$

$$= \nabla\left((Aw-y)^T(Aw-y) + \lambda w^T w\right) = \nabla\left(w^T A^T A w - 2w^T A^T y + y^T y + \lambda w^T w\right)$$

$$= 2A^T A w - 2A^T y + 2\lambda w = (2A^T A + 2\lambda I) w - 2A^T y$$

Now:

$$\|\nabla f(u) - \nabla f(v)\| = \|(2A^TA + 2\lambda I)u - 2A^T y - (2A^TA + 2\lambda I)v + 2A^T y\| =$$

$$= \|2(A^TA + \lambda I)(u-v)\| = 2\|A^TA(u-v) + \lambda(u-v)\| \underset{\text{by } \triangle \text{ inequality}}{\leq} 2(\|A^TA(u-v)\| + \lambda\|u-v\|)$$

$$\underset{\text{by our assumption}}{\leq} \overbrace{2(\lambda_{max} + \lambda)}^{L} \|u - v\| = L\|u-v\|.$$ So this shows $f(w)$ is always smooth, and the upper bound on L is $2(\lambda_{max} + \lambda)$. □

↓ for b)

b) First, let $\lambda = 0$. Then:

$\|\nabla f(u) - \nabla f(v)\| = 2\|A^TA(u-v)\| \geq 2\lambda_{min}\|u-v\|$, but λ_{min} might be 0, so this only proves $\|\nabla f(u) - \nabla f(v)\| \geq 0$, so possibly $\|\nabla f(u) - \nabla f(v)\| = 0$, so f may not be strongly convex. D

When $\lambda > 0$, we have

$\|\nabla f(u) - \nabla f(v)\| = 2\|(A^TA + \lambda I)(u-v)\|$, we know that $A^TA + \lambda I$ is a symmetric matrix, so by the properties of any symmetric matrix M, $\|Mx\| \geq \lambda_{min}^{M}\|x\|$, where λ_{min}^{M} is the smallest eigenvalue of M. So:

$\|\nabla f(u) - \nabla f(v)\| = 2\|(A^TA + \lambda I)(u-v)\| \geq 2\lambda_{min}^{A^TA+\lambda I}\|u-v\| =$
$= 2(\lambda_{min} + \lambda)\|u-v\| = d\|u-v\| \rightarrow$ so it is strongly convex)

So the lower bound on d is $2(\lambda_{min} + \lambda)$.

(3) First, note I assume $\pi(t), \pi^*$ are row vectors for all t (as per the problem setup).

Then if $\pi = \pi^* \Rightarrow \pi^* = \pi^* P \Leftrightarrow \pi^* - \pi^* P = 0$

Let: $f(\pi) = \|\pi - \pi P\|^2$ (L2 norm by assumption). Then,

$f(\pi^*) = \|\pi^* - \pi^* P\|^2 = 0$.

Also:

$$f(\pi) = \|\pi - \pi P\|^2 = \|\pi(I-P)\|^2 = \overbrace{\pi(I-P)}^{\text{row vector}}(\pi(I-P))^T =$$

$$= \pi(I-P)(I-P)^T\pi^T \Rightarrow \nabla f(\pi) = \underbrace{2\pi(I-P)(I-P)^T}_{\text{row vector}}$$

Given this setup:

a) The gradient projection algorithm is:

$$\pi(t+1) = \left[\pi(t) - \gamma 2\pi(t)(I-P)(I-P)^T\right]^+$$

where $[\pi]^+ = \underset{p \in \Pi}{\operatorname{argmin}} \|p - \pi\|$, and $\Pi = \{\pi \mid \sum_{i=1}^{N} \pi_i = 1, \pi \geq 0\}$.

For the algorithm to converge, f has to be Lipschutz smooth (assuming constant step size). So:

$\|\nabla f(\pi) - \nabla f(\alpha)\| = \|2\pi(I-P)(I-P)^T - 2\alpha(I-P)(I-P)^T\| = 2\|(\pi-\alpha)(I-P)(I-P)^T\|$

$\underset{\text{norm inequality}}{\leq} 2\|\pi - \alpha\| \underbrace{\|(I-P)(I-P)^T\|}_{\text{this is a matrix norm, defined as } \|M\| = \max_{x \neq 0} \frac{\|Mx\|}{\|x\|}} = 2\|(I-P)(I-P)^T\| \, \|\pi - \alpha\| = L\|\pi - \alpha\|$. So:

$L = 2\|(I-P)(I-P)^T\|$. Then for convergence: $\|\pi(t+1)-\pi^*\|^2 =$

$$\|\pi(t+1)-\pi(t)+\pi(t)-\pi^*\|^2 = \|\pi(t)-\pi^*\|^2 + 2\underbrace{(\pi(t+1)-\pi(t))}_{-\gamma\nabla f(\pi(t))}(\pi(t)-\pi^*)^T + \|\pi(t+1)-\pi(t)\|^2$$

$$= \|\pi(t)-\pi^*\|^2 - 2\gamma(\nabla f(\pi(t)) - \underbrace{\nabla f(\pi^*)}_{=0})(\pi(t)-\pi^*)^T + \gamma^2\|\nabla f(\pi(t))\|^2 \leq$$

by Lipschitz with L as above

$$\leq \|\pi(t)-\pi^*\|^2 + \left(\gamma^2 - 2\frac{\gamma}{L}\right)\|\nabla f(\pi(t))\|^2 \Rightarrow$$

I need $\|\pi(t)-\pi^*\|^2 - \|\pi(t+1)-\pi^*\|^2 > 0$ for convergence, which translates to $2\frac{\gamma}{L} - \gamma^2 > 0 \Rightarrow \gamma < \frac{2}{L} = \frac{2}{2\|(I-P)(I-P)^T\|} = \frac{1}{\|(I-P)(I-P)^T\|}$.

and by properties of Matrix norms $\|(I-P)(I-P)^T\| = \lambda_{max}^{(I-P)(I-P)^T} = \lambda_{max} :=$ largest eigenvalue of $(I-P)(I-P)^T$, so $\gamma < \frac{1}{\lambda_{max}}$.

6) By the hint, let Q be a matrix such that:

$$P_{i1} = \varepsilon + Q_{i1} \Rightarrow Q_{i1} = P_{i1} - \varepsilon$$

$$P_{ij} = Q_{ij} \Rightarrow Q_{ij} = P_{ij},\ j \neq 1.$$

Then: $xP = \left[\sum_i x_i(\varepsilon + Q_{i1}) \quad \sum_i x_i Q_{i2} \ \dots \ \sum x_i Q_{iN}\right] =$

$= \left[\sum_i x_i\varepsilon \quad 0 \quad 0 \dots 0\right] + xQ$

similarly:

$yP = \left[\sum_i y_i \quad 0 \quad 0 \dots 0\right] + yQ$

$$\Rightarrow (xP - yP) = \underbrace{\left[\sum x_i\varepsilon - \sum y_i\varepsilon \quad 0 \dots 0\right]}_{=\varepsilon(\sum x_i - \sum y_i) = \varepsilon(1-1) = 0} + (xQ - yQ) = xQ - yQ$$

Finally:

$$\|xP - yP\|_1 = \|xQ - yQ\|_1 = \sum_{j=1}^N \left|\sum_{i=1}^N (x_i - y_i)Q_{ij}\right| \underset{\text{by } \Delta \text{ ineq}}{\leq} \sum_{j=1}^N \sum_{i=1}^N |x_i - y_i| \overbrace{|Q_{ij}|}^{\geq 0} =$$

swap sum order and take out indep

$$= \sum_{i=1}^N |x_i - y_i| \underbrace{\sum_{j=1}^N Q_{ij}}_{=\sum P_{ij} - \varepsilon = 1-\varepsilon} = (1-\varepsilon)\sum_{i=1}^N |x_i - y_i| = (1-\varepsilon)\|x-y\|_1.$$

So this shows that $\|xP - yP\|_1 \leq (1-\varepsilon)\|x-y\|_1$, so the iteration is a contraction mapping □.

(4) a). Since f is smooth and convex, then by the growth lemma:

$$f(y) \le f(x) + [\nabla f(x)]^T (y-x) + \frac{L}{2}\|y-x\|^2$$, where $\|\cdot\|$ is the L2 norm.

Then, let $x = x(t)$, $y = x(t+1) = x(t) - \gamma \nabla f(x(t))$. Substituting:

$$f(x(t+1)) \le f(x(t)) + [\nabla f(x(t))]^T(-\gamma \nabla f(x(t))) + \frac{L}{2}\gamma^2 \|\nabla f(x(t))\|^2 =$$

$$= f(x(t)) + \|\nabla f(x(t))\|^2 \left(-\gamma + \frac{L}{2}\gamma^2\right) = f(x(t)) - \underbrace{\left(\gamma - \frac{L}{2}\gamma^2\right)}_{>0 \text{ when } \gamma < \frac{2}{L}} \|\nabla f(x(t))\|^2.$$

So this shows:

$$f(x(t+1)) \le f(x(t)) - \alpha \|\nabla f(x(t))\|^2, \text{ where } \alpha = \left(\gamma - \frac{L}{2}\gamma^2\right) > 0$$

when $\gamma < \frac{2}{L}$, i.e. the objective function is non-increasing $\square$.

b) By smoothness, we know that $\dfrac{\|\nabla f(x) - \nabla f(y)\|}{\|x-y\|} \le L$.

Now, using the same lemma as before, with $y = x(t)$, $x = x^*$:

$$\underbrace{f(x(t)) - \underbrace{f(x^*)}_{\text{moved this over}}}_{\delta(t)} \le \underbrace{[\nabla f(x^*)]^T (x(t) - x^*)}_{0} + \frac{L}{2}\|x(t) - x^*\|^2 \le 0 + \frac{\|\nabla f(x(t)) - \overbrace{\nabla f(x^*)}^{0}\|}{2\|x(t)-x^*\|}\|x(t)-x^*\|^2$$

by L ineq above

$$= \frac{1}{2}\|\nabla f(x(t))\| \cdot \|x(t) - x^*\| \le \|x(t) - x^*\| \cdot \|\nabla f(x(t))\| \quad \square$$

Then: First, by the smoothness assumption (and convexity) (and $\gamma < \frac{2}{L}$), we already know that:

$$\|x(t) - x^*\|^2 - \|x(t+1) - x^*\|^2 > 0 \Rightarrow \|x(t) - x^*\|^2 > \|x(t+1) - x^*\|^2 \Rightarrow \|x(0) - x^*\|^2 \ge \|x(t) - x^*\|^2$$

$\forall t \ge 0$. So: $-\dfrac{1}{\|x(t) - x^*\|^2} \le -\dfrac{1}{\|x(0) - x^*\|^2}$.

Next ↷

Notice that

$$\delta(t+1)-\delta(t)=f(x(t+1))-f(x(t))\underset{\substack{\text{by a)}\\ \text{proof}}}{\leq}-\alpha\|\nabla f(x(t))\|^2\leq-\frac{\alpha}{\|x(t)-x^*\|^2}\delta^2(t)\leq$$

(since $\delta^2(t)\leq\|x(t)-x^*\|^2\cdot\|\nabla f(x(t))\|^2$, so $-\frac{\delta^2(t)}{\|x(t)-x^*\|^2}\geq-\|\nabla f(x(t))\|^2$)

$$\underset{\text{by before}}{\leq}-\frac{\alpha}{\|x(0)-x^*\|^2}\delta^2(t).\iff\delta(t+1)\leq\delta(t)-\frac{\alpha}{\|x(0)-x^*\|^2}\delta^2(t)\quad\square$$

(which shows it).

c) Assume $\delta(t+1)\leq\delta(t)-w\delta^2(t)$, $w>0$

Since $w>0$ and $\delta^2(t)\geq 0$, $\delta(t+1)\leq\delta(t)-w\delta^2(t)\leq\delta(t)-0=\delta(t)$

(so $-w\delta^2(t)\leq 0$)

So $\delta(t+1)\leq\delta(t)$ and $\frac{1}{\delta(t)}\leq\frac{1}{\delta(t+1)}$

Now:

$$\delta(t+1)\leq\delta(t)-w\delta^2(t)\iff w\leq\frac{\delta(t)-\delta(t+1)}{\delta^2(t)}$$

} can divide as long as $\delta^2(t)>0$

$$\iff w\leq\frac{\delta(t)-\delta(t+1)}{\delta(t)\cdot\delta(t)}\leq\frac{\delta(t)-\delta(t+1)}{\delta(t)\delta(t+1)}=\frac{1}{\delta(t+1)}-\frac{1}{\delta(t)}$$

$$\iff\frac{1}{\delta(t+1)}-\frac{1}{\delta(t)}\geq w\quad\square$$

(which shows it).