8 Supplementary Material

8.1 Proof of Theorem 1

First, note that the update of the Proximal Point (PP) method for the bilinear problem (Assumption 1) can be written as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \mathbf{B} \mathbf{y}_{k+1}, \tag{16}$$
$$\mathbf{y}_{k+1} = \mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_{k+1}. \tag{17}$$

We can simplify the above iterations and write them as an explicit algorithm as follows:

$$\mathbf{x}_{k+1} = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k), \tag{18}$$
$$\mathbf{y}_{k+1} = (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}(\mathbf{y}_k + \eta \mathbf{B}^\top\mathbf{x}_k). \tag{19}$$

Let us define the symmetric matrices $\mathbf{Q}_x = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}$ and $\mathbf{Q}_y = (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}$. Based on these definitions, and the expressions in (18) and (19) we can show that the sum $\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2$ can be written as

$$\begin{aligned}\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 = \|\mathbf{Q}_x\mathbf{x}_k\|^2 + \eta^2\|\mathbf{Q}_x\mathbf{B}\mathbf{y}_k\|^2 + \|\mathbf{Q}_y\mathbf{y}_k\|^2 + \eta^2\|\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k\|^2 \\ - 2\eta\mathbf{x}_k^\top\mathbf{Q}_x\mathbf{B}\mathbf{y}_k + 2\eta\mathbf{y}_k^\top\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k. \end{aligned} \tag{20}$$

To simplify the expression in (20) we first prove the following lemma which is also useful in the rest of proofs.

Lemma 1. *The matrices $\mathbf{B} \in \mathbb{R}^{d\times d}$, $\mathbf{Q}_x = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}$, and $\mathbf{Q}_y = (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}$ satisfy the following properties:*

$$\mathbf{Q}_x\mathbf{B} = \mathbf{B}\mathbf{Q}_y, \tag{21}$$
$$\mathbf{Q}_y\mathbf{B}^\top = \mathbf{B}^\top\mathbf{Q}_x. \tag{22}$$

Proof. Let $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$ be the singular value decomposition of $\mathbf{B}$. Here $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices and $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{B}$ as the diagonal entries. Then, we have:

$$\begin{aligned}\mathbf{Q}_x\mathbf{B} &= (\mathbf{I} + \eta^2\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top\mathbf{V}\mathbf{\Lambda}\mathbf{U}^\top)^{-1}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top \\ &= (\mathbf{U}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})\mathbf{U}^\top)^{-1}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top \\ &= \mathbf{U}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})^{-1}\mathbf{U}^\top\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top \\ &= \mathbf{U}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})^{-1}\mathbf{\Lambda}\mathbf{V}^\top \end{aligned} \tag{23}$$

Here we used the property that $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}$. Now, we simplify the other side to get:

$$\begin{aligned}\mathbf{B}\mathbf{Q}_y &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top(\mathbf{I} + \eta^2\mathbf{V}\mathbf{\Lambda}\mathbf{U}^\top\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top)^{-1} \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top(\mathbf{V}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})\mathbf{V}^\top)^{-1} \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top\mathbf{V}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})^{-1}\mathbf{V}^\top \\ &= \mathbf{U}\mathbf{\Lambda}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})^{-1}\mathbf{V}^\top \end{aligned} \tag{24}$$

Now, since $\mathbf{U}(\eta^2\mathbf{\Lambda}^2 + \mathbf{I})^{-1}\mathbf{\Lambda}^2\mathbf{V}^\top = \mathbf{U}\mathbf{\Lambda}(\eta^2\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{V}^\top$, the claim in (21) follows. Using a similar argument we can also prove the equality in (22). □

Using the result in Lemma 1 we can show that

$$\mathbf{x}_k^\top\mathbf{Q}_x\mathbf{B}\mathbf{y}_k = \mathbf{x}_k^\top\mathbf{B}\mathbf{Q}_y\mathbf{y}_k = \mathbf{y}_k^\top\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k, \tag{25}$$

where the second equality holds as $\mathbf{a}^\top\mathbf{b} = \mathbf{b}^\top\mathbf{a}$. Hence, the expression in (26) can be simplified as

$$\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 = \|\mathbf{Q}_x\mathbf{x}_k\|^2 + \eta^2\|\mathbf{Q}_x\mathbf{B}\mathbf{y}_k\|^2 + \|\mathbf{Q}_y\mathbf{y}_k\|^2 + \eta^2\|\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k\|^2. \tag{26}$$

We simplify equation (26) as follows. Consider the term involving $\mathbf{x}_k$. We have

$$\begin{aligned}\|\mathbf{Q}_x\mathbf{x}_k\|^2 + \eta^2\|\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k\|^2 &= \mathbf{x}_k^\top\mathbf{Q}_x^2\mathbf{x}_k + \eta^2\mathbf{x}_k^\top\mathbf{B}\mathbf{Q}_y^2\mathbf{B}^\top\mathbf{x}_k \\ &= \mathbf{x}_k^\top(\mathbf{Q}_x^2 + \eta^2\mathbf{B}\mathbf{Q}_y^2\mathbf{B}^\top)\mathbf{x}_k \end{aligned} \tag{27}$$

Now we use Lemma 1 to simplify (27) as follows

$$\begin{aligned}\|\mathbf{Q}_x\mathbf{x}_k\|^2 + \eta^2\|\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k\|^2 &= \mathbf{x}_k^\top(\mathbf{Q}_x^2 + \eta^2\mathbf{B}\mathbf{Q}_y^2\mathbf{B}^\top)\mathbf{x}_k \\ &= \mathbf{x}_k^\top(\mathbf{Q}_x^2 + \eta^2\mathbf{B}\mathbf{Q}_y\mathbf{B}^\top\mathbf{Q}_x)\mathbf{x}_k \\ &= \mathbf{x}_k^\top(\mathbf{Q}_x^2 + \eta^2\mathbf{B}\mathbf{B}^\top\mathbf{Q}_x\mathbf{Q}_x)\mathbf{x}_k \\ &= \mathbf{x}_k^\top(\mathbf{I} + \eta^2\mathbf{B}\mathbf{B}^\top)\mathbf{Q}_x^2\mathbf{x}_k \\ &= \mathbf{x}_k^\top(\mathbf{I} + \eta^2\mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{x}_k, \end{aligned} \tag{28}$$

where the last equality follows by replacing $\mathbf{Q}_x$ by its definition. The same simplification follows for the terms invovling $\mathbf{y}_k$ which leads to the expression

$$\|\mathbf{Q}_y\mathbf{y}_k\|^2 + \eta^2\|\mathbf{Q}_x\mathbf{B}\mathbf{y}_k\|^2 = \mathbf{y}_k^\top(\mathbf{I} + \eta^2\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{y}_k. \tag{29}$$

Substitute $\|\mathbf{Q}_x\mathbf{x}_k\|^2 + \eta^2\|\mathbf{Q}_y\mathbf{B}^\top\mathbf{x}_k\|^2$ and $\|\mathbf{Q}_y\mathbf{y}_k\|^2 + \eta^2\|\mathbf{Q}_x\mathbf{B}\mathbf{y}_k\|^2$ in (26) with the expressions in (28) and (29), respectively, to obtain

$$\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 = \mathbf{x}_k^\top(\mathbf{I} + \eta^2\mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{x}_k + \mathbf{y}_k^\top(\mathbf{I} + \eta^2\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{y}_k. \tag{30}$$

Now, using the expression in (30) and the fact that $\lambda_{\min}(\mathbf{B}^T\mathbf{B}) = \lambda_{\min}(\mathbf{B}\mathbf{B}^T)$ we can write

$$\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 \leq \left(\frac{1}{1 + \eta^2\lambda_{\min}(\mathbf{B}^\top\mathbf{B})}\right)(\|\mathbf{x}_k\|^2 + \|\mathbf{y}_k\|^2), \tag{31}$$

and the claim in Theorem 1 follows.

8.2 Proof of Theorem 2

The update of PP method can be written as

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), \\ \mathbf{y}_{k+1} &= \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}). \end{aligned} \tag{32}$$

Consider the function $\phi_f : \mathbb{R}^m \to \mathbb{R}$ defined as

$$\phi_{\mathbf{y}_{k+1}}(\mathbf{x}) := f(\mathbf{x}, \mathbf{y}_{k+1}) + \frac{1}{2\eta}\|\mathbf{x} - \mathbf{x}_k\|^2. \tag{33}$$

It is easy to check that ϕ_f is $\mu_x + \frac{1}{\eta}$ strongly convex, and it also can be verified that $\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x}} \phi_f(\mathbf{x})$. Hence, using strong convexity of ϕ_f, for any $\mathbf{x} \in \mathbb{R}^m$, we have

$$\phi_{\mathbf{y}_{k+1}}(\mathbf{x}) - \phi_{\mathbf{y}_{k+1}}(\mathbf{x}_{k+1}) \geq \frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x} - \mathbf{x}_{k+1}\|^2, \tag{34}$$

where we used the fact that $\nabla\phi_{\mathbf{y}_{k+1}}(\mathbf{x}_{k+1}) = \mathbf{0}$. Replace $\phi_{\mathbf{y}_{k+1}}(\mathbf{x})$ and $\phi_{\mathbf{y}_{k+1}}(\mathbf{x}_{k+1})$ with their definition in (33) and further set $\mathbf{x} = \mathbf{x}^*$ to obtain

$$\begin{aligned} f(\mathbf{x}^*, \mathbf{y}_{k+1}) - f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) &\geq \frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}^*\|^2 + \frac{1}{2\eta}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &\geq \frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}^*\|^2. \end{aligned} \tag{35}$$

Once again, consider the function:

$$\phi_{\mathbf{x}_{k+1}}(\mathbf{y}) = -f(\mathbf{x}_{k+1}, \mathbf{y}) + \frac{1}{2\eta}\|\mathbf{y} - \mathbf{y}_k\|^2. \tag{36}$$

It is $\mu_y + \frac{1}{\eta}$ strongly convex and is minimized at $\mathbf{y}_{k+1}$. Therefore, for any $\mathbf{y} \in \mathbb{R}^n$, we have

$$\phi_{\mathbf{x}_{k+1}}(\mathbf{y}) - \phi_{\mathbf{x}_{k+1}}(\mathbf{y}_{k+1}) \geq \frac{1}{2}\left(\mu_y + \frac{1}{\eta}\right)\|\mathbf{y} - \mathbf{y}_{k+1}\|^2 \tag{37}$$

since $\nabla\phi_{\mathbf{x}_{k+1}}(\mathbf{y}_{k+1}) = \mathbf{0}$. Replace $\phi_{\mathbf{x}_{k+1}}(\mathbf{y})$ and $\phi_{\mathbf{x}_{k+1}}(\mathbf{y}_{k+1})$ with their definitions and further set $\mathbf{y} = \mathbf{y}^*$ to obtain

$$\begin{aligned} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - f(\mathbf{x}_{k+1}, \mathbf{y}^*) &\geq \frac{1}{2}\left(\mu_y + \frac{1}{\eta}\right)\|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 - \frac{1}{2\eta}\|\mathbf{y}_k - \mathbf{y}^*\|^2 + \frac{1}{2\eta}\|\mathbf{y}_{k+1} - \mathbf{y}_k\|^2 \\ &\geq \frac{1}{2}\left(\mu_y + \frac{1}{\eta}\right)\|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 - \frac{1}{2\eta}\|\mathbf{y}_k - \mathbf{y}^*\|^2. \end{aligned} \tag{38}$$

The saddle point property implies that the optimal solution set $(\mathbf{x}^*, \mathbf{y}^*)$ satisfies the following inequalities for any $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$:

$$f(\mathbf{x}^*, \mathbf{y}) \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq f(\mathbf{x}, \mathbf{y}^*). \tag{39}$$

In particular, by setting $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}_{k+1}, \mathbf{y}_{k+1})$ we obtain that

$$f(\mathbf{x}^*, \mathbf{y}_{k+1}) \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq f(\mathbf{x}_{k+1}, \mathbf{y}^*). \tag{40}$$

Now, considering (35), by adding and subtracting $f(\mathbf{x}^*, \mathbf{y}^*)$ we can write

$$\begin{aligned} &f(\mathbf{x}^*, \mathbf{y}_{k+1}) - f(\mathbf{x}^*, \mathbf{y}^*) + f(\mathbf{x}^*, \mathbf{y}^*) - f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \\ &\quad \geq \frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}^*\|^2, \end{aligned} \tag{41}$$

Regroup the terms to obtain

$$\begin{aligned} &f(\mathbf{x}^*, \mathbf{y}_{k+1}) - f(\mathbf{x}^*, \mathbf{y}^*) \\ &\quad \geq \frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}^*\|^2 - f(\mathbf{x}^*, \mathbf{y}^*) + f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}). \end{aligned} \tag{42}$$

By using the inequality in (40) we can write

$$\frac{1}{2}\left(\mu_x + \frac{1}{\eta}\right)\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}^*\|^2 - f(\mathbf{x}^*, \mathbf{y}^*) + f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \leq 0 \tag{43}$$

Similarly, considering (38), we can write

$$\begin{aligned} &f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - f(\mathbf{x}^*, \mathbf{y}^*) + f(\mathbf{x}^*, \mathbf{y}^*) - f(\mathbf{x}_{k+1}, \mathbf{y}^*) \\ &\quad \geq \frac{1}{2}\left(\mu_y + \frac{1}{\eta}\right)\|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 - \frac{1}{2\eta}\|\mathbf{y}_k - \mathbf{y}^*\|^2, \end{aligned} \tag{44}$$

and, therefore,

$$\frac{1}{2}\left(\mu_y + \frac{1}{\eta}\right)\|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 - \frac{1}{2\eta}\|\mathbf{y}_k - \mathbf{y}^*\|^2 - f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) + f(\mathbf{x}^*, \mathbf{y}^*) \leq 0. \tag{45}$$

Add equations (43) and (45), and use the definition $\mu = \min\{\mu_x, \mu_y\}$ to obtain

$$\frac{1}{2}\left(\mu + \frac{1}{\eta}\right)\left(\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 + \|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2\right) \leq \frac{1}{2\eta}\left(\|\mathbf{x}_k - \mathbf{x}^*\|^2 + \|\mathbf{y}_k - \mathbf{y}^*\|^2\right). \tag{46}$$

Regrouping the terms and using the definition $r_k = \|\mathbf{x}_k - \mathbf{x}^*\|^2 + \|\mathbf{y}_k - \mathbf{y}^*\|^2$ leads to

$$\begin{aligned} r_{k+1} &\leq \frac{1}{\eta}\left(\mu + \frac{1}{\eta}\right)^{-1} r_k \\ &= \frac{1}{1+\eta\mu} r_k, \end{aligned} \tag{47}$$

and the proof is complete.

8.3 Proof of Proposition 1

We start from the Proximal Point (PP) dynamics and show that an $\mathcal{O}(\eta^2)$ approximation of this dynamics leads to OGDA. The PP updates are as follows

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f\big(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\big) \\ \mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f\big(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\big) \end{aligned}$$

By writing the Taylor's expansion of $\nabla_{\mathbf{x}} f$, we obtain

$$\begin{aligned} &\nabla_{\mathbf{x}} f\left(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\right) \\ &= \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)[\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - \mathbf{x}_k] \\ &\qquad + \nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)[\mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - \mathbf{y}_k] + o(\eta) \\ &= \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) - \eta\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \\ &\qquad + \eta\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) + o(\eta). \end{aligned} \tag{48}$$

Using this expression, we have

$$\begin{aligned} \mathbf{x}_{k+1} = \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) &+ \eta^2\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \\ &- \eta^2\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) + o(\eta^2) \end{aligned} \tag{49}$$

On adding and subtracting the term $\eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k)$, we get

$$\begin{aligned} \mathbf{x}_{k+1} = \mathbf{x}_k - 2\eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) &+ \eta\big(\eta\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) + \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) \\ &- \eta\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\big) + o(\eta^2) \end{aligned} \tag{50}$$

Note that from the Taylors expansion of $\nabla_{\mathbf{xx}} f, \nabla_{\mathbf{x}} f$ and the PP updates, we have

$$\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k) = \nabla_{\mathbf{xx}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + \mathcal{O}(\eta), \quad \nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta) \tag{51}$$

which leads to

$$\eta\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \eta\nabla_{\mathbf{xx}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1})\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta^2) \tag{52}$$

Again, from the Taylor's expansion of $\nabla_{\mathbf{xy}} f, \nabla_{\mathbf{y}} f$ and the PP updates, we have

$$\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k) = \nabla_{\mathbf{xy}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + \mathcal{O}(\eta), \quad \nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta) \tag{53}$$

which implies that

$$\eta\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \eta\nabla_{\mathbf{xy}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1})\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta^2) \tag{54}$$

Making the approximations of Equations (52) and (54) in Equation (50) yields

$$\begin{aligned} \mathbf{x}_{k+1} = \mathbf{x}_k - 2\eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) &+ \eta\big(\eta\nabla_{\mathbf{xx}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1})\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) \\ &- \eta\nabla_{\mathbf{xy}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1})\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta^2)\big) + o(\eta^2) \end{aligned} \tag{55}$$

We also know that

$$\nabla_{\mathbf{x}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + \mathcal{O}(\eta^2) = \eta \nabla_{\mathbf{x}\mathbf{x}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) - \eta \nabla_{\mathbf{x}\mathbf{y}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) \nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta^2)$$

Making this substitution back in Equation (55), we get

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k - 2\eta \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta \big(\nabla_{\mathbf{x}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + \mathcal{O}(\eta^2) \big) + o(\eta^2) \\ &= \mathbf{x}_k - 2\eta \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta \nabla_{\mathbf{x}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + o(\eta^2) \end{aligned} \tag{56}$$

which is equivalent to the OGDA update plus an additional error term of order $o(\eta^2)$. The same analysis can be done for the dual updates as well to obtain

$$\mathbf{y}_{k+1} = \mathbf{y}_k + 2\eta \nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) - \eta \nabla_{\mathbf{x}} f(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) + o(\eta^2). \tag{57}$$

This shows that the OGDA updates and the PP updates differ by $o(\eta^2)$.

8.4 Proof of Theorem 3

We define the following symmetric matrices

$$\begin{aligned} \mathbf{E}_x &= \mathbf{I} - \eta^2 \mathbf{B}\mathbf{B}^\top - (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}, \\ \mathbf{E}_y &= \mathbf{I} - \eta^2 \mathbf{B}^\top\mathbf{B} - (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}. \end{aligned}$$

We rewrite the properties of $\mathbf{E}_x$ and $\mathbf{E}_y$ which are

$$\|\mathbf{E}_x\|, \|\mathbf{E}_y\| \leq \frac{\eta^4 \lambda_{\max}(\mathbf{B}^\top\mathbf{B})^2}{1 - \eta^2 \sqrt{\lambda^2_{\max}(\mathbf{B}^\top\mathbf{B})}} = e \tag{58}$$

$$\mathbf{E}_x \mathbf{B} = \mathbf{B}\mathbf{E}_y \tag{59}$$

$$\mathbf{E}_y \mathbf{B}^\top = \mathbf{B}^\top \mathbf{E}_x \tag{60}$$

Recall that the update of OGDA for the bilinear problem can be written as

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k - 2\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1}, \\ \mathbf{y}_{k+1} &= \mathbf{x}_k + 2\eta \mathbf{B}^\top \mathbf{x}_k + \eta \mathbf{B}^\top \mathbf{x}_{k-1}. \end{aligned}$$

The update for the variable $\mathbf{x}$ can be written as an approximate variant of the PP update as follows

$$\begin{aligned} \mathbf{x}_{k+1} &= (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\ &\quad - \big[\big(\mathbf{x}_k - 2\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} \big) - \big((\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B} y_k) \big) \big] \\ &= (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\ &\quad - \big[(-\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) + \mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \big] \end{aligned} \tag{61}$$

Therefore, the error between the OGDA and Proximal updates for the variable $\mathbf{x}$ is given by

$$(-\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) + \mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \tag{62}$$

We first derive an upper bound for the term in the first parentheses $(-\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k)$.

Using the OGDA update, we have:

$$\begin{aligned} -\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} &= -\eta \mathbf{B}(\mathbf{y}_k - \mathbf{y}_{k-1}) \\ &= -\eta \mathbf{B}(2\eta \mathbf{B}^\top \mathbf{x}_{k-1} - \eta \mathbf{B}^\top \mathbf{x}_{k-2}) \end{aligned} \tag{63}$$

Therefore, we can write

$$
\begin{aligned}
&(-\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) \\
&\quad = (-2\eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-2} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) \\
&\quad = (\eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-1} - \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-2} - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k)
\end{aligned}
$$

Once again, using the OGDA updates for $(\mathbf{x}_k - \mathbf{x}_{k-1})$ and $(\mathbf{x}_{k-1} - \mathbf{x}_{k-2})$, we have

$$
\begin{aligned}
&(\eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_k - \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-1} - \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-1} + \eta^2 \mathbf{B}\mathbf{B}^\top \mathbf{x}_{k-2} - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) \\
&\quad = (-2\eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_{k-1} + 3\eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_{k-2} - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_{k-3} - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k) \\
&\quad = -\eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-1})
\end{aligned} \tag{64}
$$

Therefore, considering the expressions in (62) and (64) the error between the updates of OGDA and PP for the variable $\mathbf{x}$ can be written as

$$
\begin{aligned}
&\big(\mathbf{x}_k - 2\eta \mathbf{B}\mathbf{y}_k + \eta \mathbf{B}\mathbf{y}_{k-1}\big) - \big((\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B} y_k)\big) \\
&\quad = \mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3})
\end{aligned} \tag{65}
$$

We apply the same argument for the update of the variable $\mathbf{y}$. Combining these results we obtain that the update of OGDA can be written as

$$
\begin{aligned}
\mathbf{x}_{k+1} &= (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) + \mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\
&\quad - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3}) \\
\mathbf{y}_{k+1} &= (\mathbf{I} + \eta^2 \mathbf{B}^\top \mathbf{B})^{-1}(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) \\
&\quad + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top(\mathbf{x}_k + 2\mathbf{x}_{k-1} - 3\mathbf{x}_{k-2} + \mathbf{x}_{k-3})
\end{aligned} \tag{66}
$$

As in the proof of Theorem 1, we define $\mathbf{Q}_x = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}$ and $\mathbf{Q}_y = (\mathbf{I} + \eta^2 \mathbf{B}^\top \mathbf{B})^{-1}$. Then, we can show that

$$
\begin{aligned}
\|\mathbf{x}_{k+1}\|^2 &\leq (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{Q}_x^2 (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\
&\quad + \|\mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3})\|^2 \\
&\quad + 2(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{Q}_x \big(\mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\
&\quad - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3})\big) \\
\|\mathbf{y}_{k+1}\|^2 &\leq (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{Q}_y^2 (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) \\
&\quad + \|\mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top(\mathbf{x}_k + 2\mathbf{x}_{k-1} - 3\mathbf{x}_{k-2} + \mathbf{x}_{k-3})\|^2 \\
&\quad + 2(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{Q}_y \big(\mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) \\
&\quad + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top(\mathbf{x}_k + 2\mathbf{x}_{k-1} - 3\mathbf{x}_{k-2} + \mathbf{x}_{k-3})\big)
\end{aligned} \tag{67}
$$

On summing the two sides, we have:

$$
\begin{aligned}
&\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 \\
&\leq (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{Q}_x^2 (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) + (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{Q}_y^2 (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) \\
&+ \|\mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3})\|^2 \\
&+ \|\mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top(\mathbf{x}_k + 2\mathbf{x}_{k-1} - 3\mathbf{x}_{k-2} + \mathbf{x}_{k-3})\|^2 \\
&+ 2(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{Q}_x \big(\mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}(\mathbf{y}_k + 2\mathbf{y}_{k-1} - 3\mathbf{y}_{k-2} + \mathbf{y}_{k-3})\big) \\
&+ 2(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{Q}_y \big(\mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top(\mathbf{x}_k + 2\mathbf{x}_{k-1} - 3\mathbf{x}_{k-2} + \mathbf{x}_{k-3})\big)
\end{aligned} \tag{68}
$$

Define $r_k = \|\mathbf{x}_k\|^2 + \|\mathbf{y}_k\|^2$. We have:

$$r_{k+1} \leq \max_{i\in\{k,k-1,k-2,k-3\}} \Big[\mathbf{x}_i^\top(\mathbf{Q}_x + 2\mathbf{E}_x^2 + 2\eta^2\mathbf{B}\mathbf{B}^\top\mathbf{E}_x^2 + 30\eta^6(\mathbf{B}\mathbf{B}^\top)^3 + 2\mathbf{Q}_x\mathbf{E}_x + 2\eta^2\mathbf{B}\mathbf{Q}_y\mathbf{E}_y\mathbf{B}^\top - 20\eta^3\mathbf{B}^\top\mathbf{B}\mathbf{B}^\top\mathbf{Q}_x)\mathbf{x}_i + \mathbf{y}_i^\top(\mathbf{Q}_y + 2\mathbf{E}_y^2 + 2\eta^2\mathbf{B}\mathbf{B}^\top\mathbf{E}_y^2 + 30\eta^6(\mathbf{B}\mathbf{B}^\top)^3 + 2\mathbf{Q}_y\mathbf{E}_y + 2\eta^2\mathbf{B}\mathbf{Q}_x\mathbf{E}_x\mathbf{B}^\top - 20\eta^3\mathbf{B}^\top\mathbf{B}\mathbf{B}^\top\mathbf{Q}_y)\mathbf{y}_i\Big] \tag{69}$$

And for $\eta = \frac{1}{40\sqrt{\lambda_{\max}(\mathbf{B}^\top\mathbf{B})}}$.

$$\begin{aligned} r_{k+1} &\leq \max_{i\in\{k,k-1,k-2,k-3\}} \Big[\mathbf{x}_i^\top(\mathbf{I} - \frac{1}{2}\eta^2\mathbf{B}\mathbf{B}^\top + \frac{1}{4}\eta^2\mathbf{B}\mathbf{B}^\top)\mathbf{x}_i + \mathbf{y}_i^\top(\mathbf{I} - \frac{1}{2}\eta^2\mathbf{B}^\top\mathbf{B} + \frac{1}{4}\eta^2\mathbf{B}^\top\mathbf{B})\mathbf{y}_i\Big] \\ &\leq \left(1 - \frac{1}{800\kappa}\right)\max\{r_k, r_{k-1}, r_{k-2}, r_{k-3}\} \end{aligned} \tag{70}$$

8.5 Proof of Theorem 4

We define $\mathbf{z} = [\mathbf{x};\mathbf{y}]$ and $F(\mathbf{z}) = [\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{y}); -\nabla_{\mathbf{y}} f(\mathbf{x},\mathbf{y})]$. Note that Assumption 3 gives us $\|F(\mathbf{z}_1) - F(\mathbf{z}_2)\| \leq 2L\|\mathbf{z}_1 - \mathbf{z}_2\|$. The OGDA updates can be compactly written as:

$$\mathbf{z}_{k+1} = \mathbf{z}_k - 2\eta F(\mathbf{z}_k) + \eta F(\mathbf{z}_{k-1}) \tag{71}$$

We write the update in terms of the Proximal Point method with an error $\boldsymbol{\varepsilon}_k = \eta(F(\mathbf{z}_{k+1}) - 2F(\mathbf{z}_k) + F(\mathbf{z}_{k-1}))$ as follows:

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \eta F(\mathbf{z}_{k+1}) + \boldsymbol{\varepsilon}_k \tag{72}$$

On rearranging Equation (72) and using the fact that $F(\mathbf{z}^*) = 0$, where $\mathbf{z}^* = [\mathbf{x}^*;\mathbf{y}^*]$, we get:

$$\eta(F(\mathbf{z}_{k+1}) - F(\mathbf{z}^*)) = \mathbf{z}_k - \mathbf{z}_{k+1} + \eta(F(\mathbf{z}_{k+1}) - F(\mathbf{z}_k)) - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) \tag{73}$$

On squaring Equation (73) and using Young's inequality, we get:

$$\eta^2\|F(\mathbf{z}_{k+1}) - F(\mathbf{z}^*)\|^2 \leq 3\|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2 + 3\eta^2(2L)^2\|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2 + 3\eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \tag{74}$$

Now, using strong convexity, and substituting $\eta = 1/8L$, we get:

$$\frac{\mu^2}{64L^2}\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq 4\max\{\|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2, \|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2\} \tag{75}$$

The following part of the proof is inspired by the result of Theorem 1 of Gidel et al. (2019). For OGDA iterates, we have:

$$\mathbf{z}_k - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) = \mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \eta F(\mathbf{z}_k) \tag{76}$$

On subtracting $\mathbf{z}^*$ from both sides and squaring, we have:

$$\begin{aligned} &\|\mathbf{z}_k - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 \\ &= \|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 + \eta^2\|F(\mathbf{z}_k)\|^2 - 2\eta\langle F(\mathbf{z}_k), \mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\rangle \\ &= \|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - 2\eta\langle F(\mathbf{z}_k), \mathbf{z}_k - \mathbf{z}^*\rangle - 2\langle \eta F(\mathbf{z}_k), \eta F(\mathbf{z}_{k-1})\rangle + \eta^2\|F(\mathbf{z}_k)\|^2 \\ &= \|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - 2\eta\langle F(\mathbf{z}_k), \mathbf{z}_k - \mathbf{z}^*\rangle + \eta^2\|F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})\|^2 - \eta^2\|F(\mathbf{z}_{k-1})\|^2 \\ &\leq \|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - 2\eta\langle F(\mathbf{z}_k), \mathbf{z}_k - \mathbf{z}^*\rangle + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 - \eta^2\|F(\mathbf{z}_{k-1})\|^2 \end{aligned} \tag{77}$$

However, since:

$$\langle F(\mathbf{z}_k), \mathbf{z}_k - \mathbf{z}^* \rangle \geq \mu \|\mathbf{z}_k - \mathbf{z}^*\|^2 \tag{78}$$

and using Young's inequality we have

$$\|\mathbf{z}_k - \mathbf{z}^*\|^2 \leq \frac{1}{2}\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - \|\eta F(\mathbf{z}_{k-1})\|^2 \tag{79}$$

Substituting Equations (78) and (79) in Equation (77), we have:

$$\begin{aligned}\eta\mu & \left(\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - 2\eta^2\|F(\mathbf{z}_{k-1})\|^2\right) \\ & \leq \|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 - \|\mathbf{z}_k - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 \\ & \qquad + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 - \eta^2\|F(\mathbf{z}_{k-1})\|^2\end{aligned} \tag{80}$$

which on rearranging gives:

$$\begin{aligned}\|\mathbf{z}_k & - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 \\ & \leq (1-\eta\mu)\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \\ & \qquad - \eta^2(1-2\eta\mu)\|F(\mathbf{z}_{k-1})\|^2\end{aligned} \tag{81}$$

However, for the OGDA iterates:

$$\begin{aligned}\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 &= \eta^2\|F(\mathbf{z}_{k-1} + F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})\|^2 \\ &\leq 2\eta^2\|F(\mathbf{z}_{k-1}\|^2 + 2\eta^2\|F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})\|^2 \\ &\leq 2\eta^2\|F(\mathbf{z}_{k-1})\|^2 + 2\eta^2(2L)^2\|\mathbf{z}_{k-1} - \mathbf{z}_{k-2}\|^2\end{aligned} \tag{82}$$

which can be written as:

$$\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \leq 4\eta^2\|F(\mathbf{z}_{k-1})\|^2 + 4\eta^2(2L)^2\|\mathbf{z}_{k-1} - \mathbf{z}_{k-2}\|^2 - \|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \tag{83}$$

Substituting Equation (83) in Equation (81), we get:

$$\begin{aligned}\|\mathbf{z}_k & - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 + \eta^2 L^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \\ & \leq (1-\eta\mu)\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 + 4\eta^4(2L)^4\|\mathbf{z}_{k-1} - \mathbf{z}_{k-2}\|^2 \\ & \qquad - \eta^2(1 - 2\eta\mu - 4\eta^2(2L)^2)\|F(\mathbf{z}_{k-1})\|^2\end{aligned} \tag{84}$$

For $\eta \leq 1/8L$, we have: $1 - 2\eta\mu - 4\eta^2(2L)^2 > 0$ and therefore, can ignore the last term, which gives us (for $\eta \leq 1/8L$)

$$\begin{aligned}\|\mathbf{z}_k & - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \\ & \leq (1-\eta\mu)\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 + 4\eta^4(2L)^4\|\mathbf{z}_{k-1} - \mathbf{z}_{k-2}\|^2\end{aligned} \tag{85}$$

since $\eta \leq 1/8L$, we have $(1-\eta\mu) \geq 4\eta^2(2L)^2$, which gives:

$$\begin{aligned}\|\mathbf{z}_k & - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \\ & \leq (1-\eta\mu)\left(\|\mathbf{z}_{k-1} - \eta(F(\mathbf{z}_{k-1}) - F(\mathbf{z}_{k-2})) - \mathbf{z}^*\|^2 + \eta^2(2L)^2\|\mathbf{z}_{k-1} - \mathbf{z}_{k-2}\|^2\right)\end{aligned} \tag{86}$$

which gives us:

$$\|\mathbf{z}_k - \eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})) - \mathbf{z}^*\|^2 + \eta^2(2L)^2\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \leq (1-\eta\mu)^k\left(\|\mathbf{z}_0 - \mathbf{z}^*\|^2\right) \tag{87}$$

in particular, for $\eta = \frac{1}{8L}$:

$$\|\mathbf{z}_k - \mathbf{z}_{k-1}\|^2 \leq 64\left(1 - \frac{\mu}{8L}\right)^k\left(\|\mathbf{z}_0 - \mathbf{z}^*\|^2\right) \tag{88}$$

Substituting Equation (88) back in Equation (75), we get:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq \left(1 - \frac{\mu}{8L}\right)^k \times (16384\kappa^2\|\mathbf{z}_0 - \mathbf{z}^*\|^2) \tag{89}$$

On defining $\hat{r}_0 = 16384\kappa^2\|\mathbf{z}_0 - \mathbf{z}^*\|^2$, we have:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq \left(1 - \frac{\mu}{8L}\right)^k \hat{r}_0. \tag{90}$$

8.6 Proof of Theorem 5

The generalized OGDA method for bilinear problems is given by:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\alpha+\beta)\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1}$$
$$\mathbf{y}_{k+1} = \mathbf{y}_k + (\alpha+\beta)\mathbf{B}\mathbf{y}_k - \beta\mathbf{B}\mathbf{y}_{k-1}$$

We compare this with the Proximal Point (PP) method with stepsize α. The proof follows long the exact same lines as the proof of Theorem 3.

We define the following symmetric matrices

$$\mathbf{E}_x = \mathbf{I} - \alpha^2\mathbf{B}\mathbf{B}^\top - (\mathbf{I} + \alpha^2\mathbf{B}\mathbf{B}^\top)^{-1},$$
$$\mathbf{E}_y = \mathbf{I} - \alpha^2\mathbf{B}^\top\mathbf{B} - (\mathbf{I} + \alpha^2\mathbf{B}^\top\mathbf{B})^{-1}.$$

We rewrite the properties of $\mathbf{E}_x$ and $\mathbf{E}_y$ which are

$$\|\mathbf{E}_x\|, \|\mathbf{E}_y\| \leq \frac{\alpha^4\lambda_{\max}(\mathbf{B}^\top\mathbf{B})^2}{1-\alpha^2\sqrt{\lambda^2_{\max}(\mathbf{B}^\top\mathbf{B})}} = e \tag{91}$$

$$\mathbf{E}_x\mathbf{B} = \mathbf{B}\mathbf{E}_y \tag{92}$$
$$\mathbf{E}_y\mathbf{B}^\top = \mathbf{B}^\top\mathbf{E}_x \tag{93}$$

Therefore, the error between the OGDA and Proximal updates for the variable $\mathbf{x}$ is given by

$$(-\beta\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1} + \alpha^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k) + \mathbf{E}_x(\mathbf{x}_k - \alpha\mathbf{B}\mathbf{y}_k) \tag{94}$$

We first derive an upper bound for the term in the first parentheses $(-\beta\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1} + \alpha^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k)$.

Using the generalized OGDA update, we have:

$$\begin{aligned}-\beta\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1} &= -\beta\mathbf{B}(\mathbf{y}_k - \mathbf{y}_{k-1})\\ &= -\beta\mathbf{B}((\alpha+\beta)\mathbf{B}^\top\mathbf{x}_{k-1} - \beta\mathbf{B}^\top\mathbf{x}_{k-2})\end{aligned} \tag{95}$$

Therefore, we can write

$$\begin{aligned}&(-\beta\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1} + \alpha^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k)\\ &\quad= (\alpha^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha\beta\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-1} - \beta^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-1} + \beta^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-2} - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k)\end{aligned}$$

Once again, using the generalized OGDA updates for $(\mathbf{x}_k - \mathbf{x}_{k-1})$ and $(\mathbf{x}_{k-1} - \mathbf{x}_{k-2})$, we have

$$\begin{aligned}&(\alpha^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha\beta\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-1} - \beta^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-1} + \beta^2\mathbf{B}\mathbf{B}^\top\mathbf{x}_{k-2} - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k)\\ &\quad= \alpha(\alpha-\beta)\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha\beta(\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-1} + \alpha\beta^2\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-2}\\ &\qquad+ \beta^2(\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-2} - \beta^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-3} - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k\\ &\quad= \alpha(\alpha-\beta)\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha\beta(\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-1} + \beta^2(2\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-2}\\ &\qquad- \beta^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-3} - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k\end{aligned} \tag{96}$$

Therefore, considering the expressions in (94) and (96) the error between the updates of OGDA and PP for the variable $\mathbf{x}$ can be written as

$$\begin{aligned}&\big(\mathbf{x}_k - (\alpha+\beta)\mathbf{B}\mathbf{y}_k + \beta\mathbf{B}\mathbf{y}_{k-1}\big) - \big((\mathbf{I} + \alpha^2\mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \alpha\mathbf{B}y_k)\big)\\ &\quad= \mathbf{E}_x(\mathbf{x}_k - \alpha\mathbf{B}\mathbf{y}_k) + \alpha(\alpha-\beta)\mathbf{B}\mathbf{B}^\top\mathbf{x}_k - \alpha\beta(\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-1}\\ &\qquad+ \beta^2(2\alpha+\beta)\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-2} - \beta^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_{k-3} - \alpha^3\mathbf{B}\mathbf{B}^\top\mathbf{B}\mathbf{y}_k\end{aligned} \tag{97}$$

Now, the convergence proof follows along the same lines as the proof of Theorem 3. We set $\eta = \max\{\alpha, \beta\}$, and we need the additional assumption:

$$|\alpha - \beta| \leq \mathcal{O}(\eta^3/\alpha) \tag{98}$$

due to the presence of the term $\alpha(\alpha-\beta)\mathbf{B}\mathbf{B}^\top \mathbf{x}_k$, Let

$$\alpha - K\alpha^2 \leq \beta \leq \alpha \tag{99}$$

On making these substitutions, we get the same result as Theorem 3.

8.7 Proof of Proposition 2

The Extragradient updates can be written as

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k)) \\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k))
\end{aligned}$$

By writing the Taylor's expansion of $\nabla_{\mathbf{x}} f$ we obtain that

$$\begin{aligned}
&\nabla_{\mathbf{x}} f\left(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k)\right) \\
&= \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)[\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) - \mathbf{x}_k] \\
&\qquad + \nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)[\mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) - \mathbf{y}_k] + o(\eta) \\
&= \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) - \eta\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + o(\eta).
\end{aligned} \tag{100}$$

Use this expression to write

$$\begin{aligned}
\mathbf{x}_{k+1} = \mathbf{x}_k &- \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta^2\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) \\
&- \eta^2\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + o(\eta^2).
\end{aligned} \tag{101}$$

By following the same argument for $\mathbf{y}$ we obtain

$$\begin{aligned}
\mathbf{y}_{k+1} = \mathbf{y}_k &+ \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta^2\nabla_{\mathbf{yy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) \\
&- \eta^2\nabla_{\mathbf{yx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + o(\eta^2)
\end{aligned} \tag{102}$$

Now we find a second order approximation for the Proximal Point Method. Note that the update of the proximal point method for variable $\mathbf{x}$ can be written as

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \\
&= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f\big(\mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), \mathbf{y}_k + \eta\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\big)
\end{aligned} \tag{103}$$

where in the second equality we replaced $\mathbf{x}_{k+1}$ and $\mathbf{y}_{k+1}$ in the gradient with their updates. Hence, using Taylor's series we can show that

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta^2\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) \\
&\quad - \eta^2\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) + o(\eta^2) \\
&= \mathbf{x}_k - \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta^2\nabla_{\mathbf{xx}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) \\
&\quad - \eta^2\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + o(\eta^2),
\end{aligned} \tag{104}$$

where in the second equality we used the fact that $\nabla_{\mathbf{x}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta)$ and $\nabla_{\mathbf{y}} f(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) + \mathcal{O}(\eta)$. Similarly, we find the approximation of the update of $\mathbf{y}$ which leads to

$$\begin{aligned}
\mathbf{y}_{k+1} = \mathbf{y}_k &+ \eta\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + \eta^2\nabla_{\mathbf{yy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}_k) \\
&- \eta^2\nabla_{\mathbf{xy}} f(\mathbf{x}_k, \mathbf{y}_k)\nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}_k) + o(\eta^2).
\end{aligned} \tag{105}$$

Comparing the expressions in (101) and (102) with the ones in (104) and (105) implies that the difference between the updates of PP and EG is at most $o(\eta^2)$ and this completes the proof.

8.8 Proof of Theorem 6

Define the following symmetric error matrices

$$\mathbf{E}_x = \mathbf{I} - \eta^2 \mathbf{B}\mathbf{B}^\top - (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}, \qquad \mathbf{E}_y = \mathbf{I} - \eta^2 \mathbf{B}^\top\mathbf{B} - (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1} \tag{106}$$

which are useful to characterize the difference between the updates of EG and PP for a bilinear problem. Note that we can bound the norms of $\mathbf{E}_x$ and $\mathbf{E}_y$ as

$$\begin{aligned}\|\mathbf{E}_x\| &\leq \eta^4 \sqrt{\lambda_{\max}^4(\mathbf{B}\mathbf{B}^\top)} + \eta^6 \sqrt{\lambda_{\max}^6(\mathbf{B}\mathbf{B}^\top)} + \cdots \\ &= \frac{\eta^4 \lambda_{\max}(\mathbf{B}\mathbf{B}^\top)^2}{1 - \eta^2 \sqrt{\lambda_{\max}^2(\mathbf{B}\mathbf{B}^\top)}},\end{aligned} \tag{107}$$

and similarly

$$\|\mathbf{E}_y\| \leq \frac{\eta^4 \lambda_{\max}(\mathbf{B}^\top\mathbf{B})^2}{1 - \eta^2 \sqrt{\lambda_{\max}^2(\mathbf{B}^\top\mathbf{B})}}. \tag{108}$$

Since $\lambda_{\max}(\mathbf{B}^\top\mathbf{B}) = \lambda_{\max}(\mathbf{B}\mathbf{B}^\top)$, we have:

$$\|\mathbf{E}_x\|, \|\mathbf{E}_y\| \leq \frac{\eta^4 \lambda_{\max}(\mathbf{B}^\top\mathbf{B})^2}{1 - \eta^2 \sqrt{\lambda_{\max}^2(\mathbf{B}^\top\mathbf{B})}} := e \tag{109}$$

Also, from Lemma 1 in the proof of Theorem 1, and the definitions of the error matrices in (106) it can be verified that

$$\mathbf{E}_x \mathbf{B} = \mathbf{B}\mathbf{E}_y \tag{110}$$

$$\mathbf{E}_y \mathbf{B}^\top = \mathbf{B}^\top \mathbf{E}_x \tag{111}$$

Moreover, using the definitions of $\mathbf{E}_x$ and $\mathbf{E}_y$ in (106), the EG updates can be written as

$$\mathbf{x}_{k+1} = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) - \eta^3 \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k + \mathbf{E}_x(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k), \tag{112}$$

$$\mathbf{y}_{k+1} = (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \eta^3 \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_k + \mathbf{E}_y(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k). \tag{113}$$

As in the proof of Theorem 1, we define $\mathbf{Q}_x = (\mathbf{I} + \eta^2 \mathbf{B}\mathbf{B}^\top)^{-1}$ and $\mathbf{Q}_y = (\mathbf{I} + \eta^2 \mathbf{B}^\top\mathbf{B})^{-1}$. Using these definitions we can show that

$$\begin{aligned}\|\mathbf{x}_{k+1}\|^2 = &(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{Q}_x^2 (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) + \eta^6 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k \\ &+ (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{E}_x^2 (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) + 2(\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{E}_x \mathbf{Q}_x (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k) \\ &- 2\eta^3 (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)^\top \mathbf{E}_x \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k - 2\eta^3 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{Q}_x (\mathbf{x}_k - \eta \mathbf{B}\mathbf{y}_k)\end{aligned} \tag{114}$$

$$\begin{aligned}\|\mathbf{y}_{k+1}\|^2 = &(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{Q}_y^2 (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + \eta^6 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_k + (\mathbf{y}_k \\ &+ \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{E}_y^2 (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) + 2(\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{E}_y \mathbf{Q}_y (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k) \\ &+ 2\eta^3 (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)^\top \mathbf{E}_y \mathbf{B}^\top \mathbf{B}\mathbf{B}^\top \mathbf{x}_k + 2\eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{Q}_y (\mathbf{y}_k + \eta \mathbf{B}^\top \mathbf{x}_k)\end{aligned} \tag{115}$$

Now before adding the two sides of the expressions in (114) and (115), note that some of the cross terms in (114) and (115) cancel out. For instance, using Lemma 1 and Equations (110) and (111) we can show that

$$\begin{aligned}-\eta^3 \mathbf{x}_k^\top \mathbf{E}_x \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{y}_k + \eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{y}_k &= -\eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{E}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k + \eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{y}_k \\ &= -\eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{E}_x \mathbf{B}\mathbf{y}_k + \eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{y}_k \\ &= -\eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{y}_k + \eta^3 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{y}_k \\ &= 0\end{aligned}$$

By using similar arguments it can be shown that summing two sides of the expressions in (114) and (115) leads to

$$\begin{aligned}
&\|\mathbf{x}_{k+1}\|^2 + \|\mathbf{y}_{k+1}\|^2 \\
&= \mathbf{x}_k^\top \mathbf{Q}_x^2 \mathbf{x}_k + \eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{Q}_x^2 \mathbf{B}\mathbf{y}_k + \eta^6 \mathbf{y}_k^\top (\mathbf{B}^\top\mathbf{B})^3 \mathbf{y}_k + \mathbf{x}_k^\top \mathbf{E}_x^2 \mathbf{x}_k + \eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{E}_x^2 \mathbf{B}\mathbf{y}_k \\
&\quad + 2\mathbf{x}_k^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{x}_k + 2\eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{B}\mathbf{y}_k + 2\eta^4 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k + 2\eta^4 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{Q}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k \\
&\quad + \mathbf{y}_k^\top \mathbf{Q}_y^2 \mathbf{y}_k + \eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{Q}_y^2 \mathbf{B}^\top \mathbf{x}_k + \eta^6 \mathbf{x}_k^\top (\mathbf{B}\mathbf{B}^\top)^3 \mathbf{x}_k + \mathbf{y}_k^\top \mathbf{E}_y^2 \mathbf{y}_k + \eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{E}_y^2 \mathbf{B}^\top \mathbf{x}_k \\
&\quad + 2\mathbf{y}_k^\top \mathbf{E}_y \mathbf{Q}_y \mathbf{y}_k + 2\eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{E}_y \mathbf{Q}_y \mathbf{B}^\top \mathbf{x}_k + 2\eta^4 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{E}_x \mathbf{B}\mathbf{B}^\top \mathbf{x}_k + 2\eta^4 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{Q}_x \mathbf{B}\mathbf{B}^\top \mathbf{x}_k \\
&= \mathbf{x}_k^\top \mathbf{Q}_x \mathbf{x}_k + \eta^6 \mathbf{y}_k^\top (\mathbf{B}^\top\mathbf{B})^3 \mathbf{y}_k + \mathbf{x}_k^\top \mathbf{E}_x^2 \mathbf{x}_k + \eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{E}_x^2 \mathbf{B}\mathbf{y}_k \\
&\quad + 2\mathbf{x}_k^\top \mathbf{E}_x \mathbf{x}_k + 2\eta^4 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k + 2\eta^4 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{Q}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k \\
&\quad + \mathbf{y}_k^\top \mathbf{Q}_y \mathbf{y}_k + \eta^6 \mathbf{x}_k^\top (\mathbf{B}\mathbf{B}^\top)^3 \mathbf{x}_k + \mathbf{y}_k^\top \mathbf{E}_y^2 \mathbf{y}_k + \eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{E}_y^2 \mathbf{B}^\top \mathbf{x}_k \\
&\quad + 2\mathbf{y}_k^\top \mathbf{E}_y \mathbf{y}_k + 2\eta^4 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{E}_x \mathbf{B}\mathbf{B}^\top \mathbf{x}_k + 2\eta^4 \mathbf{x}_k^\top \mathbf{B}\mathbf{B}^\top \mathbf{Q}_x \mathbf{B}\mathbf{B}^\top \mathbf{x}_k
\end{aligned} \tag{116}$$

where in the second equality we used the simplifications

$$\begin{aligned}
\mathbf{x}_k^\top \mathbf{Q}_x^2 \mathbf{x}_k + \eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{Q}_y^2 \mathbf{B}^\top \mathbf{x}_k &= \mathbf{x}_k^\top \mathbf{Q}_x^2 \mathbf{x}_k + \eta^2 \mathbf{x}_k^\top \mathbf{Q}_x^2 \mathbf{B}\mathbf{B}^\top \mathbf{Q}_x \mathbf{x}_k = \mathbf{x}_k^\top \mathbf{Q}_x \mathbf{x}_k \\
\mathbf{y}_k^\top \mathbf{Q}_y^2 \mathbf{y}_k + \eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{Q}_x^2 \mathbf{B}\mathbf{y}_k &= \mathbf{y}_k^\top \mathbf{Q}_y^2 \mathbf{y}_k + \eta^2 \mathbf{y}_k^\top \mathbf{Q}_y^2 \mathbf{B}^\top \mathbf{B}\mathbf{Q}_y \mathbf{y}_k = \mathbf{y}_k^\top \mathbf{Q}_y \mathbf{y}_k
\end{aligned} \tag{117}$$

as well as

$$\begin{aligned}
\mathbf{x}_k^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{x}_k + \eta^2 \mathbf{x}_k^\top \mathbf{B}\mathbf{E}_y \mathbf{Q}_y \mathbf{B}^\top \mathbf{x}_k &= \mathbf{x}_k^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{x}_k + \eta^2 \mathbf{x}_k^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{B}\mathbf{B}^\top \mathbf{x}_k = \mathbf{x}_k^\top \mathbf{E}_x \mathbf{x}_k \\
\mathbf{y}_k^\top \mathbf{E}_y \mathbf{Q}_y \mathbf{y}_k + \eta^2 \mathbf{y}_k^\top \mathbf{B}^\top \mathbf{E}_x \mathbf{Q}_x \mathbf{B}\mathbf{y}_k &= \mathbf{y}_k^\top \mathbf{E}_y \mathbf{Q}_y \mathbf{y}_k + \eta^2 \mathbf{y}_k^\top \mathbf{E}_y \mathbf{Q}_y \mathbf{B}^\top \mathbf{B}\mathbf{y}_k = \mathbf{y}_k^\top \mathbf{E}_y \mathbf{y}_k
\end{aligned} \tag{118}$$

Define $r_k = \|\mathbf{x}_k\|^2 + \|\mathbf{y}_k\|^2$. We have:

$$\begin{aligned}
&r_{k+1} \leq \\
&\mathbf{x}_k^\top (\mathbf{Q}_x + 2\mathbf{E}_x + \mathbf{E}_x^2 + \eta^6 (\mathbf{B}\mathbf{B}^\top)^3 + \eta^2 \mathbf{B}\mathbf{E}_y^2 \mathbf{B}^\top + 2\eta^4 \mathbf{B}\mathbf{B}^\top \mathbf{E}_x \mathbf{B}\mathbf{B}^\top + 2\eta^4 \mathbf{B}\mathbf{B}^\top \mathbf{Q}_x \mathbf{B}\mathbf{B}^\top)\mathbf{x}_k \\
&+ \mathbf{y}_k^\top (\mathbf{Q}_y + 2\mathbf{E}_y + \mathbf{E}_y^2 + \eta^6 (\mathbf{B}^\top\mathbf{B})^3 + \eta^2 \mathbf{B}^\top \mathbf{E}_x^2 \mathbf{B} + 2\eta^4 \mathbf{B}^\top \mathbf{B}\mathbf{E}_y \mathbf{B}^\top \mathbf{B} + 2\eta^4 \mathbf{B}^\top \mathbf{B}\mathbf{Q}_y \mathbf{B}^\top \mathbf{B})\mathbf{y}_k
\end{aligned} \tag{119}$$

Choosing $\eta = \frac{1}{2\sqrt{2\lambda_{\max}(\mathbf{B}^\top\mathbf{B})}}$, we have:

$$\begin{aligned}
r_{k+1} &\leq \mathbf{x}_k^\top (\mathbf{I} - \frac{1}{2}\eta^2 \mathbf{B}\mathbf{B}^\top + \frac{1}{4}\eta^2 \mathbf{B}\mathbf{B}^\top)\mathbf{x}_k + \mathbf{y}_k^\top (\mathbf{I} - \frac{1}{2}\eta^2 \mathbf{B}^\top\mathbf{B} + \frac{1}{4}\eta^2 \mathbf{B}^\top\mathbf{B})\mathbf{y}_k \\
&\leq \left(1 - \frac{1}{20\kappa}\right) r_k
\end{aligned} \tag{120}$$

8.9 Proof of Theorem 7

Define $\mathbf{z} = [\mathbf{x}; \mathbf{y}]$ and $F(\mathbf{z}) = [\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}); -\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})]$. Note that Assumption 3 gives us $\|F(\mathbf{z}_1) - F(\mathbf{z}_2)\| \leq 2L\|\mathbf{z}_1 - \mathbf{z}_2\|$. The EG updates can be compactly written as:

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \eta F(\mathbf{z}_{k+1/2}) \tag{121}$$

where

$$\mathbf{z}_{k+1/2} = \mathbf{z}_k - \eta F(\mathbf{z}_k) \tag{122}$$

We write the update in terms of the Proximal Point method with an error $\boldsymbol{\varepsilon}_k = \eta(F(\mathbf{z}_{k+1}) - F(\mathbf{z}_{k+1/2}))$ as follows:

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \eta F(\mathbf{z}_{k+1}) + \boldsymbol{\varepsilon}_k \tag{123}$$

On squaring and simplifying this expression, we have:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 = \|\mathbf{z}_k - \mathbf{z}^*\|^2 - \|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2 - 2\eta(F(\mathbf{z}_{k+1}) + \boldsymbol{\varepsilon}_k)^\top (\mathbf{z}_{k+1} - \mathbf{z}^*) \tag{124}$$

where $\mathbf{z}^* = [\mathbf{x}^*; \mathbf{y}^*]$ (Note that $r_k = \|\mathbf{z}_k - \mathbf{z}^*\|^2$). We simplify the right hand side of Equation (124) as follows-

$$\begin{aligned}
&\|\mathbf{z}_k - \mathbf{z}^*\|^2 - \|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2 - 2\eta(F(\mathbf{z}_{k+1}) + \boldsymbol{\varepsilon}_k)^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) \\
&= \|\mathbf{z}_k - \mathbf{z}^*\|^2 - \|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2 - 2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) \\
&= \|\mathbf{z}_k - \mathbf{z}^*\|^2 - 2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) - \|\mathbf{z}_{k+1} - \mathbf{z}_{k+1/2} + \mathbf{z}_{k+1/2} - \mathbf{z}_k\|^2 \\
&= \|\mathbf{z}_k - \mathbf{z}^*\|^2 - 2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) - \|\mathbf{z}_{k+1} - \mathbf{z}_{k+1/2}\|^2 - \|\mathbf{z}_{k+1/2} - \mathbf{z}_k\|^2 \\
&\qquad - 2(\mathbf{z}_{k+1} - \mathbf{z}_{k+1/2})^\top(\mathbf{z}_{k+1/2} - \mathbf{z}_k) \\
&= \|\mathbf{z}_k - \mathbf{z}^*\|^2 - 2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) - \|\mathbf{z}_{k+1} - \mathbf{z}_{k+1/2}\|^2 - \|\mathbf{z}_{k+1/2} - \mathbf{z}_k\|^2 \\
&\qquad - 2\eta(F(\mathbf{z}_k))^\top(\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1})
\end{aligned} \tag{125}$$

The following part of the proof is inspired by the result of Theorem 1 of Gidel et al. (2019). We simplify the inner products and give a lower bound using strong convexity as follows:

$$\begin{aligned}
&2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) + 2\eta(F(\mathbf{z}_k))^\top(\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1}) \\
&\quad = 2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1/2} - \mathbf{z}^*) + 2\eta(F(\mathbf{z}_k) - F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1}) \\
&\quad \geq 2\eta\mu\|\mathbf{z}_{k+1/2} - \mathbf{z}^*\| - 4\eta L\|\mathbf{z}_k - \mathbf{z}_{k+1/2}\|\|\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1}\|
\end{aligned} \tag{126}$$

since $F(\mathbf{z}^*) = 0$. Now, using Young's inequality, we have:

$$\begin{aligned}
&2\eta(F(\mathbf{z}_{k+1/2}))^\top(\mathbf{z}_{k+1} - \mathbf{z}^*) + 2\eta(F(\mathbf{z}_k))^\top(\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1}) \\
&\quad \geq 2\eta\mu\|\mathbf{z}_{k+1/2} - \mathbf{z}^*\|^2 - (4\eta^2 L^2\|\mathbf{z}_k - \mathbf{z}_{k+1/2}\|^2 + \|\mathbf{z}_{k+1/2} - \mathbf{z}_{k+1}\|^2)
\end{aligned} \tag{127}$$

Substituting the above inequality in Equation (125), we have:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq \|\mathbf{z}_k - \mathbf{z}^*\|^2 - 2\eta\mu\|\mathbf{z}_{k+1/2} - \mathbf{z}^*\|^2 + (4\eta^2 L^2 - 1)\|\mathbf{z}_k - \mathbf{z}_{k+1/2}\| \tag{128}$$

Since $\|\mathbf{z}_{k+1/2} - \mathbf{z}^*\|^2 \leq 2\|\mathbf{z}_k - \mathbf{z}^*\|^2 + 2\|\mathbf{z}_{k+1/2} - \mathbf{z}_k\|^2$, we have:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq (1 - \eta\mu)\|\mathbf{z}_k - \mathbf{z}^*\|^2 + (4\eta^2 L^2 + 2\eta\mu - 1)\|\mathbf{z}_k - \mathbf{z}_{k+1/2}\| \tag{129}$$

For $\eta = 1/8L$, we have $4\eta^2 L^2 + 2\eta\mu - 1 < 1$ (since $\mu \leq L$), which gives:

$$\|\mathbf{z}_{k+1} - \mathbf{z}^*\|^2 \leq \left(1 - \frac{1}{8\kappa}\right)\|\mathbf{z}_k - \mathbf{z}^*\|^2 \tag{130}$$

where $\kappa = \frac{L}{\mu}$.