A Technical Proofs

Proof of Theorem 1. Define the line interpolation between two points,

$$\theta(x) = x\theta_t + (1-x)\theta_* \ ,$$
$$\omega(x) = x\omega_t + (1-x)\omega_* \ .$$

Then the SGA dynamics can be written as (using Taylor's theorem with remainder)

$$\begin{aligned}
\begin{bmatrix}\theta_{t+1}-\theta_* \\ \omega_{t+1}-\omega_*\end{bmatrix} &= \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} - \eta \begin{bmatrix}\nabla_\theta U(\theta_t,\omega_t) \\ -\nabla_\omega U(\theta_t,\omega_t)\end{bmatrix} \ , \\
&= \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} - \eta \int_0^1 \begin{bmatrix}\nabla_{\theta\theta} U(\theta(x),\omega(x)) & \nabla_{\theta\omega} U(\theta(x),\omega(x)) \\ -\nabla_{\omega\theta} U(\theta(x),\omega(x)) & -\nabla_{\omega\omega} U(\theta(x),\omega(x))\end{bmatrix} dx \cdot \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} \ , \\
&= \int_0^1 \left(I - \eta \begin{bmatrix}\nabla_{\theta\theta} U(\theta(x),\omega(x)) & \nabla_{\theta\omega} U(\theta(x),\omega(x)) \\ -\nabla_{\omega\theta} U(\theta(x),\omega(x)) & -\nabla_{\omega\omega} U(\theta(x),\omega(x))\end{bmatrix} \right) dx \cdot \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} \ .
\end{aligned}$$

Assume that one can prove for some $r > 0$, and $(\theta,\omega) \in B_2((\theta_*,\omega_*), r)$, with a proper choice of η, the largest singular value is bounded above by 1,

$$\left\| I - \eta \begin{bmatrix}\nabla_{\theta\theta} U(\theta,\omega) & \nabla_{\theta\omega} U(\theta,\omega) \\ -\nabla_{\omega\theta} U(\theta,\omega) & -\nabla_{\omega\omega} U(\theta,\omega)\end{bmatrix} \right\|_{\text{op}} < 1 \ .$$

Then due to convexity of the operator norm, the dynamics of SGA is contracting locally because,

$$\begin{aligned}
\left\| \begin{bmatrix}\theta_{t+1}-\theta_* \\ \omega_{t+1}-\omega_*\end{bmatrix} \right\| &\leqslant \left\| \int_0^1 \left(I - \eta \begin{bmatrix}\nabla_{\theta\theta} U(\theta(x),\omega(x)) & \nabla_{\theta\omega} U(\theta(x),\omega(x)) \\ -\nabla_{\omega\theta} U(\theta(x),\omega(x)) & -\nabla_{\omega\omega} U(\theta(x),\omega(x))\end{bmatrix} \right) dx \right\|_{\text{op}} \cdot \left\| \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} \right\| \ , \\
&\leqslant \int_0^1 \left\| \left(I - \eta \begin{bmatrix}\nabla_{\theta\theta} U(\theta(x),\omega(x)) & \nabla_{\theta\omega} U(\theta(x),\omega(x)) \\ -\nabla_{\omega\theta} U(\theta(x),\omega(x)) & -\nabla_{\omega\omega} U(\theta(x),\omega(x))\end{bmatrix} \right) \right\|_{\text{op}} dx \cdot \left\| \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} \right\| \ , \\
&< \left\| \begin{bmatrix}\theta_t-\theta_* \\ \omega_t-\omega_*\end{bmatrix} \right\| \ .
\end{aligned}$$

Let's analyze the singular values of

$$I - \eta \begin{bmatrix}\nabla_{\theta\theta} U(\theta,\omega) & \nabla_{\theta\omega} U(\theta,\omega) \\ -\nabla_{\omega\theta} U(\theta,\omega) & -\nabla_{\omega\omega} U(\theta,\omega)\end{bmatrix} \ ,$$

assuming $\nabla_{\theta\theta} U(\theta,\omega) \succ 0$, $-\nabla_{\omega\omega} U(\theta,\omega) \succ 0$. Abbreviate

$$\begin{bmatrix}\nabla_{\theta\theta} U(\theta,\omega) & \nabla_{\theta\omega} U(\theta,\omega) \\ -\nabla_{\omega\theta} U(\theta,\omega) & -\nabla_{\omega\omega} U(\theta,\omega)\end{bmatrix} := \begin{bmatrix}A & C \\ -C^T & B\end{bmatrix} \ .$$

The largest singular value of

$$I - \eta \begin{bmatrix}A & C \\ -C^T & B\end{bmatrix} \ ,$$

is the square root of the largest eigenvalue of the following symmetric matrix

$$\begin{bmatrix}I-\eta A & -\eta C \\ \eta C^T & I-\eta B\end{bmatrix} \begin{bmatrix}I-\eta A & \eta C \\ -\eta C^T & I-\eta B\end{bmatrix} = \begin{bmatrix}(I-\eta A)^2 + \eta^2 CC^T & -\eta^2(AC-CB) \\ -\eta^2(C^TA - BC^T) & (I-\eta B)^2 + \eta^2 C^TC\end{bmatrix} \ .$$

It is clear that when $\eta = 0$, the largest eigenvalue of the above matrix is 1. Observe

$$\begin{aligned}
&\begin{bmatrix}(I-\eta A)^2 + \eta^2 CC^T & -\eta^2(AC-CB) \\ -\eta^2(C^TA - BC^T) & (I-\eta B)^2 + \eta^2 C^TC\end{bmatrix} = I - 2\eta \begin{bmatrix}A & 0 \\ 0 & B\end{bmatrix} + \eta^2 \begin{bmatrix}A^2 + CC^T & -AC+CB \\ -C^TA + BC^T & B^2 + C^TC\end{bmatrix} \ , \\
&\preceq \left[1 - 2\eta\lambda_{\min}\left(\begin{bmatrix}A & 0 \\ 0 & B\end{bmatrix}\right) + \eta^2\lambda_{\max}\left(\begin{bmatrix}A^2 + CC^T & -AC+CB \\ -C^TA + BC^T & B^2 + C^TC\end{bmatrix}\right)\right] I \ .
\end{aligned}$$

If we choose η to be

$$\eta = \frac{\min_{(\theta,\omega)\in B_2((\theta_*,\omega_*),r)} \lambda_{\min}\left(\begin{bmatrix} A_{\theta,\omega} & 0 \\ 0 & B_{\theta,\omega}\end{bmatrix}\right)}{\max_{(\theta,\omega)\in B_2((\theta_*,\omega_*),r)} \lambda_{\max}\left(\begin{bmatrix} A_{\theta,\omega}^2 + C_{\theta,\omega}C_{\theta,\omega}^T & -A_{\theta,\omega}C_{\theta,\omega} + C_{\theta,\omega}B_{\theta,\omega} \\ -C_{\theta,\omega}^T A_{\theta,\omega} + B_{\theta,\omega}C_{\theta,\omega}^T & B_{\theta,\omega}^2 + C_{\theta,\omega}^T C_{\theta,\omega}\end{bmatrix}\right)} = \frac{\sqrt{\alpha}}{\beta} \ ,$$

then we find

$$\begin{bmatrix} (I-\eta A)^2 + \eta^2 CC^T & -\eta^2(AC - CB) \\ -\eta^2(C^T A - BC^T) & (I-\eta B)^2 + \eta^2 C^T C \end{bmatrix} \prec \left(1 - \frac{\alpha}{\beta}\right) I \ .$$

In this case,

$$\begin{aligned} \left\|\begin{bmatrix}\theta_{t+1} - \theta_* \\ \omega_{t+1} - \omega_*\end{bmatrix}\right\| &\leqslant \sup_{(\theta,\omega)\in B_2((\theta_*,\omega_*),r)} \left\| I - \eta \begin{bmatrix} \nabla_{\theta\theta}U(\theta,\omega) & \nabla_{\theta\omega}U(\theta,\omega) \\ -\nabla_{\omega\theta}U(\theta,\omega) & -\nabla_{\omega\omega}U(\theta,\omega)\end{bmatrix}\right\|_{\mathrm{op}} \left\|\begin{bmatrix}\theta_t - \theta_* \\ \omega_t - \omega_*\end{bmatrix}\right\| \ , \\ &\leqslant \sqrt{1 - \frac{\alpha}{\beta}} \cdot \left\|\begin{bmatrix}\theta_t - \theta_* \\ \omega_t - \omega_*\end{bmatrix}\right\| \ . \end{aligned}$$

Therefore, to obtain an ϵ-minimizer one requires a number of steps equal to

$$2\frac{\beta}{\alpha}\log\frac{r}{\epsilon} \ .$$

□

Proof of Remark 1. Consider $U(\theta) = \frac{1}{2}\theta^T A\theta$ where $A \succ 0$ is strictly positive. Then gradient descent corresponds to $\theta_{t+1} = (I - \eta A)\theta_t$ and thus $\|\theta_{t+1}\| \leqslant \|I - \eta A\|_{\mathrm{op}}\|\theta_t\|$. Setting $\eta = 1/\lambda_{\max}(A)$ we have $I - \eta A \succeq 0$ so $\|I - \eta A\|_{\mathrm{op}} = \lambda_{\max}(I - \eta A) = 1 - \lambda_{\min}(A)/\lambda_{\max}(A)$. Therefore $\|\theta_t\| \leqslant \|\theta_0\|\big[1 - \lambda_{\min}(A)/\lambda_{\max}(A)\big]^t \leqslant \|\theta_0\| e^{-t\lambda_{\min}(A)/\lambda_{\max}(A)}$. The number of iterations required to obtain an ϵ-minimizer is thus bounded as $T \geqslant \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}\log\frac{r}{\epsilon}$.

□

Proof of Corollary 1. We have

$$\begin{aligned} \begin{bmatrix}\theta_{t+1} \\ \omega_{t+1}\end{bmatrix} &= \begin{bmatrix}\theta_t \\ \omega_t\end{bmatrix} - \eta \begin{bmatrix} \nabla_\theta U(\theta_t,\omega_t) \\ -\nabla_\omega U(\theta_t,\omega_t)\end{bmatrix} \ , \\ &= \left(I - \eta\begin{bmatrix} I & C \\ -C^T & I\end{bmatrix}\right)\cdot\begin{bmatrix}\theta_t \\ \omega_t\end{bmatrix} \ . \end{aligned}$$

If we define $D = \mathrm{diag}\left((1-\eta)^2 I + \eta^2 CC^T, (1-\eta)^2 I + \eta^2 C^T C\right)$ then using the Rayleigh quotient representation of $\lambda_{\min}(D)$ we obtain,

$$\left\|\begin{bmatrix}\theta_{t+1} \\ \omega_{t+1}\end{bmatrix}\right\|^2 = \begin{bmatrix}\theta_t & \omega_t\end{bmatrix} D \begin{bmatrix}\theta_t \\ \omega_t\end{bmatrix} \geqslant \lambda_{\min}(D)\left\|\begin{bmatrix}\theta_t \\ \omega_t\end{bmatrix}\right\|^2 \ .$$

On the other hand,

$$\lambda_{\min}(D) = \lambda_{\min}\left((1-\eta)^2 I + \eta^2 CC^T\right) = 1 - 2\eta + \left[1 + \lambda_{\min}(CC^T)\right]\eta^2 \geqslant \frac{\lambda_{\min}(CC^T)}{1 + \lambda_{\min}(CC^T)}$$

regardless of the choice of η, which proves the claim. □

Proof of Theorem 3. Recall that the OMD dynamics iteratively updates

$$\begin{bmatrix}\theta_{t+1} \\ \omega_{t+1}\end{bmatrix} = \left(I - 2\eta\begin{bmatrix} 0 & C \\ -C^T & 0\end{bmatrix}\right)\cdot\begin{bmatrix}\theta_t \\ \omega_t\end{bmatrix} + \eta\begin{bmatrix} 0 & C \\ -C^T & 0\end{bmatrix}\cdot\begin{bmatrix}\theta_{t-1} \\ \omega_{t-1}\end{bmatrix} \ .$$

Define the following matrices

$$R_1 = \frac{\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right) + \left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)^{1/2}}{2} \ , \tag{11}$$

$$R_2 = \frac{\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right) - \left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)^{1/2}}{2} \ . \tag{12}$$

It is easy to verify that

$$R_1 + R_2 = \left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right) \ ,$$

$$R_1 R_2 = R_2 R_1 = \frac{\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right)^2 - \left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)}{4} = -\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix} \ .$$

The commutative property $R_1 R_2 = R_2 R_1$ follows from a singular value decomposition argument: Letting $C = UDV^T$ be the SVD of C (D diagonal) one finds,

$$C\left(I - 4\eta^2 C^T C\right)^{1/2} = UD\left(I - 4\eta^2 D^2\right)^{1/2} V^T = U\left(I - 4\eta^2 D^2\right)^{1/2} DV^T = \left(I - 4\eta^2 CC^T\right)^{1/2} C \ .$$

Using the above equality, the commutative property follows

$$\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right)\left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)^{1/2} = \left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)^{1/2}\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right) \ ,$$

$$\implies \quad R_1 R_2 = R_2 R_1 \ .$$

Now we have the following relations for OMD,

$$\begin{bmatrix} \theta_{t+1} \\ \omega_{t+1} \end{bmatrix} - R_1 \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} = R_2 \left(\begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} - R_1 \begin{bmatrix} \theta_{t-1} \\ \omega_{t-1} \end{bmatrix}\right) \ ,$$

$$\begin{bmatrix} \theta_{t+1} \\ \omega_{t+1} \end{bmatrix} - R_2 \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} = R_1 \left(\begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} - R_2 \begin{bmatrix} \theta_{t-1} \\ \omega_{t-1} \end{bmatrix}\right) \ .$$

Hence

$$(R_1 - R_2)\begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} = R_1^t \left(\begin{bmatrix} \theta_1 \\ \omega_1 \end{bmatrix} - R_2 \begin{bmatrix} \theta_0 \\ \omega_0 \end{bmatrix}\right) - R_2^t \left(\begin{bmatrix} \theta_1 \\ \omega_1 \end{bmatrix} - R_1 \begin{bmatrix} \theta_0 \\ \omega_0 \end{bmatrix}\right) \ . \tag{13}$$

Let's analyze the singular values of R_1 and R_2. We have,

$$\begin{aligned}
R_1 &= \frac{\left(I - 2\eta \begin{bmatrix} 0 & C \\ -C^T & 0 \end{bmatrix}\right) + \left(I - 4\eta^2 \begin{bmatrix} CC^T & 0 \\ 0 & C^TC \end{bmatrix}\right)^{1/2}}{2} \ , \\
&= \begin{bmatrix} \frac{I + (I - 4\eta^2 CC^T)^{1/2}}{2} & -\eta C \\ \eta C^T & \frac{I + (I - 4\eta^2 C^TC)^{1/2}}{2} \end{bmatrix} \ , \\
R_1 R_1^T &= \begin{bmatrix} \frac{I + (I - 4\eta^2 CC^T)^{1/2}}{2} & -\eta C \\ \eta C^T & \frac{I + (I - 4\eta^2 C^TC)^{1/2}}{2} \end{bmatrix} \begin{bmatrix} \frac{I + (I - 4\eta^2 CC^T)^{1/2}}{2} & \eta C \\ -\eta C^T & \frac{I + (I - 4\eta^2 C^TC)^{1/2}}{2} \end{bmatrix} \ , \\
&= \begin{bmatrix} \left(\frac{I + (I - 4\eta^2 CC^T)^{1/2}}{2}\right)^2 + \eta^2 CC^T & 0 \\ 0 & \left(\frac{I + (I - 4\eta^2 C^TC)^{1/2}}{2}\right)^2 + \eta^2 C^TC \end{bmatrix} \ , \\
&= \begin{bmatrix} \frac{I + (I - 4\eta^2 CC^T)^{1/2}}{2} & 0 \\ 0 & \frac{I + (I - 4\eta^2 C^TC)^{1/2}}{2} \end{bmatrix} \ .
\end{aligned}$$

Similarly

$$R_2 R_2^T = \begin{bmatrix} \frac{I-(I-4\eta^2 CC^T)^{1/2}}{2} & 0 \\ 0 & \frac{I-(I-4\eta^2 C^T C)^{1/2}}{2} \end{bmatrix} .$$

For η small enough, the spectral radius satisfies the strict inequality $\|R_1\|_{\mathrm{op}} < 1$. Concretely, for example,

$$\begin{aligned} &\eta = \frac{1}{2\sqrt{2\lambda_{\max}(CC^T)}} \implies \\ &\frac{I+(I-4\eta^2 CC^T)^{1/2}}{2} \preceq \frac{I+(I-2\eta^2 CC^T)}{2} = I - \eta^2 CC^T \preceq \left[1 - \frac{1}{8}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right] I \ , \\ &\frac{I-(I-4\eta^2 CC^T)^{1/2}}{2} \preceq \frac{1}{2} I \ . \end{aligned}$$

Therefore $\|R_1\|_{\mathrm{op}} \leqslant \sqrt{1 - \frac{1}{8}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}}$, $\|R_2\|_{\mathrm{op}} \leqslant \sqrt{1-\frac{1}{2}}$.

The RHS of Eqn. (13) is upper bounded because

$$\begin{aligned} \left\| R_2^t \left(\begin{bmatrix} \theta_1 \\ \omega_1 \end{bmatrix} - R_1 \begin{bmatrix} \theta_0 \\ \omega_0 \end{bmatrix} \right) \right\| &\leqslant \left(1 - \frac{1}{2}\right)^{t/2} \left(\|(\theta_1, \omega_1)\| + \|(\theta_0, \omega_0)\| \right) \ , \\ \left\| R_1^t \left(\begin{bmatrix} \theta_1 \\ \omega_1 \end{bmatrix} - R_2 \begin{bmatrix} \theta_0 \\ \omega_0 \end{bmatrix} \right) \right\| &\leqslant \left(1 - \frac{1}{8}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right)^{t/2} \left(\|(\theta_1, \omega_1)\| + \|(\theta_0, \omega_0)\| \right) \ . \end{aligned}$$

Moreover the LHS satisfies

$$\left\| (R_1 - R_2) \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \right\| \geqslant \left\| \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \right\| \sqrt{\frac{1}{2}} \ .$$

Combining these inequalities we obtain

$$\|(\theta_t, \omega_t)\| \sqrt{\frac{1}{2}} \leqslant \left[\left(1 - \frac{1}{2}\right)^{t/2} + \left(1 - \frac{1}{8}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right)^{t/2} \right] \left(\|(\theta_0, \omega_0)\| + \|(\theta_1, \omega_1)\| \right) \ . \tag{14}$$

By our assumption $\|(\theta_0, \omega_0)\|, \|(\theta_1, \omega_1)\| \leqslant r$ so

$$\begin{aligned} \|(\theta_t, \omega_t)\| &\leqslant 2\sqrt{2} r \left[\left(1 - \frac{1}{2}\right)^{t/2} + \left(1 - \frac{1}{8}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right)^{t/2} \right] \ , \quad (15) \\ &\leqslant 2\sqrt{2} r \left[\exp\left(-\frac{t}{4}\right) + \exp\left(-\frac{t}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right) \right] \ , \quad (16) \end{aligned}$$

because $\forall (x, \alpha) \in \mathbb{R} \times \mathbb{R}^+ : (1-x)^\alpha \leqslant e^{-\alpha x}$. Thus

$$\|(\theta_t, \omega_t)\| \leqslant 2\sqrt{2} r \exp\left(-\frac{t}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right) \left[1 + \exp\left(-\frac{t}{4} + \frac{t}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right) \right] \ . \tag{17}$$

But $16 \geqslant 4\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)} \implies -\frac{1}{4} + \frac{1}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)} \leqslant 0$ so

$$\begin{aligned} \|(\theta_t, \omega_t)\| &\leqslant 2\sqrt{2} r \exp\left(-\frac{t}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right) [1+1] \ , \quad (18) \\ &= 4\sqrt{2} r \exp\left(-\frac{t}{16}\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}\right) \ . \quad (19) \end{aligned}$$

To sum up, when

$$T \geqslant \left\lceil 16 \frac{\lambda_{\max}(CC^T)}{\lambda_{\min}(CC^T)} \log \frac{4\sqrt{2} r}{\epsilon} \right\rceil \ ,$$

one can ensure $\|(\theta_T, \omega_T)\| \leqslant \epsilon$. □

Proof of Theorem 6. In this case, the consensus optimization satisfies the following update

$$\begin{bmatrix}\theta_{t+1}\\ \omega_{t+1}\end{bmatrix} = \left(I - \eta\begin{bmatrix}\gamma CC^T & C\\ -C^T & \gamma C^T C\end{bmatrix}\right)\cdot\begin{bmatrix}\theta_t\\ \omega_t\end{bmatrix} \ . \tag{20}$$

Again let's analyze the singular values of the operator $K := \left(I - \eta\begin{bmatrix}\gamma CC^T & C\\ -C^T & \gamma C^T C\end{bmatrix}\right)$, or equivalently, the eigenvalues of KK^T,

$$\begin{aligned}KK^T &= \begin{bmatrix}I-\eta\gamma CC^T & -\eta C\\ \eta C^T & I-\eta\gamma C^T C\end{bmatrix}\begin{bmatrix}I-\eta\gamma CC^T & \eta C\\ -\eta C^T & I-\eta\gamma C^T C\end{bmatrix} \ ,\\ &= \begin{bmatrix}(I-\eta\gamma CC^T)^2+\eta^2 CC^T & (I-\eta\gamma CC^T)\eta C - \eta C(I-\eta\gamma C^T C)\\ \eta C^T(I-\eta\gamma CC^T) - (I-\eta\gamma C^T C)\eta C^T & (I-\eta\gamma C^T C)^2+\eta^2 C^T C\end{bmatrix} \ ,\\ &= \begin{bmatrix}(I-\eta\gamma CC^T)^2+\eta^2 CC^T & 0\\ 0 & (I-\eta\gamma C^T C)^2+\eta^2 C^T C\end{bmatrix} \ .\end{aligned}$$

Now consider the largest eigenvalue of $(I-\eta\gamma CC^T)^2+\eta^2 CC^T$, for a fixed γ, with a properly chosen η. Using the SVD $C = UDV^T$, we obtain

$$\begin{aligned}&(I-\eta\gamma CC^T)^2+\eta^2 CC^T = U\left[(I-\eta\gamma D^2)^2+\eta^2 D^2\right]U^T\\ &\preceq \left[1-2\gamma\lambda_{\min}(CC^T)\eta + (\gamma^2\lambda_{\max}^2(CC^T)+\lambda_{\max}(CC^T)\eta^2\right] I \ ,\\ &= \left[1-\frac{\gamma^2\lambda_{\min}^2(CC^T)}{\gamma^2\lambda_{\max}^2(CC^T)+\lambda_{\max}(CC^T)}\right] I \ ,\end{aligned}$$

with

$$\eta = \frac{\gamma\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)+\gamma^2\lambda_{\max}^2(CC^T)} \ .$$

□

Proof of Theorem 5. In this case, the implicit update satisfies the update rule

$$\begin{bmatrix}\theta_{t+1}\\ \omega_{t+1}\end{bmatrix} = \begin{bmatrix}\theta_t\\ \omega_t\end{bmatrix} - \eta\begin{bmatrix}0 & C\\ -C^T & 0\end{bmatrix}\cdot\begin{bmatrix}\theta_{t+1}\\ \omega_{t+1}\end{bmatrix} \tag{21}$$

$$\left(I + \eta\begin{bmatrix}0 & C\\ -C^T & 0\end{bmatrix}\right)\cdot\begin{bmatrix}\theta_{t+1}\\ \omega_{t+1}\end{bmatrix} = \begin{bmatrix}\theta_t\\ \omega_t\end{bmatrix} \ . \tag{22}$$

Let's analyze the singular values of the matrix $K := \left(I + \eta\begin{bmatrix}0 & C\\ -C^T & 0\end{bmatrix}\right)$, or equivalently the root of eigenvalues of KK^T

$$\begin{aligned}KK^T &= \begin{bmatrix}I & \eta C\\ -\eta C^T & I\end{bmatrix}\begin{bmatrix}I & -\eta C\\ \eta C^T & I\end{bmatrix}\\ &= \begin{bmatrix}I+\eta^2 CC^T & 0\\ 0 & I+\eta^2 C^T C\end{bmatrix} \ .\end{aligned}$$

It is clear that the singular values of K, denoted by $\sigma_i(K)$ is sandwiched between

$$\sqrt{1+\eta^2\lambda_{\min}(CC^T)} \leqslant \sigma_i(K) \leqslant \sqrt{1+\eta^2\lambda_{\max}(CC^T)} \ .$$

If we choose $\eta = \frac{1}{\sqrt{\lambda_{\max}(CC^T)}}$, then

$$0 \leqslant \eta^2\lambda_{\min}(CC^T) \leqslant \frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)} \leqslant 1 \ .$$

Using the fact that for all $0 \leqslant t \leqslant 1$, $1 + (\sqrt{2}-1)t \leqslant \sqrt{1+t}$, then

$$\sigma_{\min}(K) \geqslant \sqrt{1+\eta^2\lambda_{\min}(CC^T)} \geqslant 1 + (\sqrt{2}-1)\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)} \ .$$

Recall Eqn. (22), using the fact for all $0 \leqslant t \leqslant 1$, $1/(1+(\sqrt{2}-1)t) \leqslant 1-(1-1/\sqrt{2})t$, we know

$$\begin{aligned}
\sigma_{\min}(K) \cdot \left\| \begin{bmatrix} \theta_{t+1} \\ \omega_{t+1} \end{bmatrix} \right\| &\leqslant \left\| \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \right\| \\
\left\| \begin{bmatrix} \theta_{t+1} \\ \omega_{t+1} \end{bmatrix} \right\| &\leqslant \frac{1}{1+(\sqrt{2}-1)\frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)}} \left\| \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \right\| \\
&\leqslant \left(1 - (1 - \frac{1}{\sqrt{2}}) \frac{\lambda_{\min}(CC^T)}{\lambda_{\max}(CC^T)} \right) \left\| \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \right\| \ .
\end{aligned}$$

To sum up, when

$$T \geqslant \left\lceil (2+\sqrt{2}) \frac{\lambda_{\max}(CC^T)}{\lambda_{\min}(CC^T)} \log \frac{r}{\epsilon} \right\rceil \ ,$$

one can ensure $\|(\theta_T, \omega_T)\| \leqslant \epsilon$. □

Proof of Theorem 4. In the simple bi-linear game case,

$$\begin{aligned}
\begin{bmatrix} \theta_{t+1} \\ \omega_{t+1} \end{bmatrix} &= \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} - \begin{bmatrix} 0 & \eta C \\ -\eta C^T & 0 \end{bmatrix} \begin{bmatrix} \theta_{t+1/2} \\ \omega_{t+1/2} \end{bmatrix} \ , \\
&= \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} - \begin{bmatrix} 0 & \eta C \\ -\eta C^T & 0 \end{bmatrix} \begin{bmatrix} I & -\gamma C \\ \gamma C^T & I \end{bmatrix} \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \ , \\
&= \begin{bmatrix} I - \eta\gamma CC^T & -\eta C \\ \eta C^T & I - \eta\gamma C^T C \end{bmatrix} \begin{bmatrix} \theta_t \\ \omega_t \end{bmatrix} \ .
\end{aligned}$$

Note this linear system is the same as that in Thm. 6. Therefore the convergence analysis follows in the same way as Thm. 6. □