A. Solution to the Matrix Riccati Differential Equation

Here we prove Proposition 1.

First, we note that a general matrix Riccati differential equation takes the form

$$\dot{P}(t) = AP(t) + P(t)A^T - P(t)RP(t) + Q, \qquad P(0) = P_0, \tag{40}$$

where $P(t) \in \mathbb{R}^{n\times n}$, and $A, R, Q, P_0 \in \mathbb{R}^{n\times n}$ are constant matrices. Associated to (40) one has the linear system

$$\begin{bmatrix} \dot{X}_1(t) \\ \dot{X}_2(t) \end{bmatrix} = \begin{bmatrix} -A^T & R \\ Q & A \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix}, \qquad \begin{bmatrix} X_1(0) \\ X_2(0) \end{bmatrix} = \begin{bmatrix} I_n \\ P_0 \end{bmatrix}. \tag{41}$$

The closed form solution of (40) follows as a consequence of lemma 6 below.

Lemma 6 (Sasagawa (1982)). *Consider the two initial value problems* (40) *and* (41). *We have:*

- *The initial value problem* (40) *has a solution in the interval* $[0, t_1]$ *if and only if the matrix* $X_1(t)$ *in the solution of the linear differential equation* (41) *is invertible for all* $t \in [0, t_1)$. *Moreover, the solution to* (40) *is unique and given by*

$$P(t) = X_2(t)X_1(t)^{-1}. \tag{42}$$

- *Let* $\bar{P}$ *be a solution to the algebraic Riccati equation (ARE)*

$$AP + PA^T - PRP + Q = 0. \tag{43}$$

Then the solution of (41) *is given by* (44) *below, where* $\tilde{A} = A - \bar{P}R$ *and* $\hat{A} = A^T - R\bar{P}$*:*

$$\begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} e^{-t\hat{A}} + \left(\int_0^t ds\, e^{-(t-s)\hat{A}} R\, e^{s\tilde{A}}\right)(P_0 - \bar{P}) \\ \bar{P}e^{-t\hat{A}} + \bar{P}\left(\int_0^t ds\, e^{-(t-s)\hat{A}} R\, e^{s\tilde{A}}\right)(P_0 - \bar{P}) + e^{t\tilde{A}}(P_0 - \bar{P}) \end{bmatrix}. \tag{44}$$

Now, we apply Lemma 6 to the Riccati equation induced by a symmetric factorization, namely,

$$\dot{X}(t) = 2X(t)Y + 2YX(t) - 4X(t)^2. \tag{45}$$

The associated linear system is

$$\begin{bmatrix} \dot{X}_1(t) \\ \dot{X}_2(t) \end{bmatrix} = \begin{bmatrix} -2Y & 4I \\ 0 & 2Y \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix}, \qquad \begin{bmatrix} X_1(0) \\ X_2(0) \end{bmatrix} = \begin{bmatrix} I_n \\ X_0 \end{bmatrix}. \tag{46}$$

and the algebraic Riccati equation is given by

$$XY + YX - 2X^2 = 0. \tag{47}$$

This equation admits the trivial solution $X = 0$. Thus, one can verify that as long as the given matrix Y is invertible we have

$$\begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} e^{-2tY} + e^{-2tY}Y^{-1}(e^{4Yt} - I)X_0 \\ e^{2tY}X_0 \end{bmatrix}. \tag{48}$$

Therefore, the unique solution to (5) is explicitly given by

$$X(t) = X_2(t)X_1(t)^{-1} = e^{2tY}X_0\left(I + Y^{-1}(e^{4tY} - I)X_0\right)^{-1} e^{2tY}, \tag{49}$$

as long as the matrix between parenthesis is invertible. Note that $Y^{-1}(e^{4tY} - I)$ is always positive semidefinite for $t \geq 0$. Thus, for all $X_0 \succeq 0$, the matrix $I + Y^{-1}(e^{4tY} - I)X_0$ is invertible and the solution of the Riccati equation is well defined for all $t \geq 0$.

B. Rate of Convergence in the Symmetric Case

In this section, we show that the solution of the Riccati equation $\dot{X}(t) = 2YX(t) + 2X(t)Y - 4X(t)^2$ converges exponentially to $\tilde{Y}$ at a rate $O(e^{-4t\sigma_{\min}(Y)})$, where $\sigma_{\min}(Y) = \min_i |\sigma_i|$ is the smallest eigenvalue of Y in magnitude. We already know that if $X(t)$ converges to X_*, then

- X_* is positive semidefinite if $X_0 \succeq 0$ and $\mathrm{rank}(X_*) \leq \mathrm{rank}(X_0)$,
- X_* is a solution to the algebraic Riccati equation $2YX + 2XY - 4X^2 = 0$.

Our proof of proposition 2 is inspired by the proofs in Callier et al. (1992; 1994); Molinari (1977) which studied the solution of the Riccati equation under a more general setting. The strategy will be to show that the algebraic Riccati equation has a unique PSD solution X_+ such that the eigenvalues of $\tilde{A} = \hat{A} = 2Y - 4X_+$ have negative real parts. Such solution is usually referred to as the strong solution or stabilizing solution in optimal control literature because it is the only solution of the algebraic Riccati equation such that the matrix $\tilde{A}$ is exponentially stable, i.e. $\exp(t\tilde{A})$ converges to 0. The stability of the matrix $\tilde{A}$ is important because it appears in the solution of the Riccati equation as can be observed in (44).

We start by proving that $\tilde{Y}$ is a solution to the algebraic Riccati equation, i.e. it is a critical point of the problem.

Due to the symmetric and positive semidefinite nature of the matrices X and Y, in our case the algebraic equation (47) can be reduced to

$$X(X - Y) = 0. \tag{50}$$

For $X = \tilde{Y}$, we thus have

$$\begin{aligned}\tilde{Y}(\tilde{Y} - Y) &= \left(\sum_{i=1}^{m} \max\{\sigma_i, 0\}\phi_i\phi_i^T\right)\left(\sum_{i=1}^{m}(\max\{\sigma_i, 0\} - \sigma_i)\phi_i\phi_i^T\right) \\ &= \left(\sum_{i}^{m} \max\{\sigma_i, 0\}\phi_i\phi_i^T\right)\left(\sum_{i=1}^{m} \min\{\sigma_i, 0\}\phi_i\phi_i^T\right).\end{aligned} \tag{51}$$

The first sum contains only the positive eigenvalues while the second sum contains only the negative ones. Therefore, no eigenvalue will appear in both sums and using the orthogonality of the vectors ϕ_i we conclude that $\tilde{Y}(\tilde{Y} - Y) = 0$.

Next, we prove that $\bar{P} = \tilde{Y}$ is the unique symmetric positive semidefinite solution of the algebraic Riccati equation such that the eigenvalues of $\tilde{A} = A - R\bar{P}$ have negative real parts. Note that $\tilde{A} = 2Y - 4\tilde{Y} = -2\hat{Y}$ where $\hat{Y} = \sum_{i=1}^{m} |\sigma_i|\phi_i\phi_i^T \succ 0$.

We proceed by contradiction. Let X_2 be a PSD solution of the (ARE) such that the eigenvalues of $\tilde{A}_2 = 2Y - 4X_2$ have negative real parts and $X_2 \neq \tilde{Y}$.

Now consider $\Delta = \tilde{Y} - X_2$. By a straightforward calculation, we can show that Δ is a solution of

$$\tilde{A}\Delta + \Delta\tilde{A} + 4\Delta^2 = 0. \tag{52}$$

Since Δ is not necessarily invertible, we consider a basis such that

$$\Delta = Z \begin{bmatrix} 0 & 0 \\ 0 & D \end{bmatrix} Z^T = Z\tilde{\Delta}Z^T, \tag{53}$$

where D is invertible. Note that our proof holds and is more trivial if Δ is invertible. We write

$$\tilde{A} = Z \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix} Z^T = Z\bar{A}Z^T \tag{54}$$

After the change of basis, equation (52) becomes

$$\bar{A}\tilde{\Delta} + \tilde{\Delta}\bar{A} + 4\tilde{\Delta}^2 = 0. \tag{55}$$

From (52) , we can deduce the following for the block matrices:

$$W_2 = 0,$$
$$W_3 = 0,$$
$$DW_4 + W_4 D + 4D^2 = 0.$$

Note that the last equation is similar to equation (52) for the invertible block D. Moreover, in the new basis, $\tilde{A}$ is a block diagonal matrix. Using the change of variable $T = D^{-1}$, we obtain the Lyapunov equation

$$TW_4 + W_4 T + 4I = 0. \tag{56}$$

Since $\tilde{A} = -2\hat{Y} \prec 0$ is invertible, the block W_4 is also invertible and its eigenvalues are a subset of the eigenvalues of $-2\hat{Y}$. As a result, the Lyapunov equation has the unique trivial solution $T = -2W_4^{-1} \succ 0$, therefore $D = -\frac{1}{2}W_4 \succ 0$.

Using the solution of (52), we can obtain a new derivation for $\tilde{A}_2$ as follows

$$\begin{aligned} \tilde{A}_2 &= 2Y - 4X_2 = 2Y - 4\tilde{Y} + 4\Delta \\ &= \tilde{A} + 4\Delta = Z \begin{bmatrix} W_1 & 0 \\ 0 & -W_4 \end{bmatrix} Z^T \end{aligned}$$

with $W_1 \prec 0$ and $-W_4 \succ 0$. This contradicts with the initial assumption on the eigenvalues of $\tilde{A}_2$ having negative real parts.

Now we can use Lemma 6 with $\bar{P} = \tilde{Y}$ to obtain a new expression of the solution. We have $\tilde{A} = \hat{A} = -2\hat{Y}$ and

$$\begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} e^{-2t\hat{Y}}\left[I + \hat{Y}^{-1}(I - e^{-4\hat{Y}t})(X_0 - \tilde{Y})\right] \\ \tilde{Y}e^{-2t\hat{Y}}\left[I + \hat{Y}^{-1}(I - e^{-4\hat{Y}t})(X_0 - \tilde{Y})\right] + e^{2t\hat{Y}}(X_0 - \tilde{Y}) \end{bmatrix}. \tag{57}$$

Therefore, the solution of the Riccati equation is also given by

$$X(t) = X_2(t)X_1(t)^{-1} = \tilde{Y} + e^{-2t\hat{Y}}(X_0 - \tilde{Y})\left[I + \hat{Y}^{-1}(I - e^{-4\hat{Y}t})(X_0 - \tilde{Y})\right]^{-1} e^{-2t\hat{Y}} \tag{58}$$

Note that the inverse exists for $t \geq 0$ when $X_0 \succeq 0$ because we have proven the existence of the solution using the previous expression (49) (see lemma 2 in Sasagawa (1982) for a detailed proof).

We introduce the function $H(t) = (X_0 - \tilde{Y})\left[I + \hat{Y}^{-1}(I - e^{-4\hat{Y}t})(X_0 - \tilde{Y})\right]^{-1}$. Thus,

$$X(t) - \tilde{Y} = e^{-2t\hat{Y}} H(t) e^{-2t\hat{Y}} \tag{59}$$

The function H has the following properties for $t \geq 0$:

- It is decreasing:

$$\frac{dH(t)}{dt} = -4H(t)e^{-4\hat{Y}t}H(t) \leq 0 \tag{60}$$

- If $I + \hat{Y}(X_0 - \tilde{Y})$ is invertible then

$$\lim_{t\to\infty} H(t) = (X_0 - \tilde{Y})[I + \hat{Y}(X_0 - \tilde{Y})]^{-1} = \tilde{H}. \tag{61}$$

- H is bounded on $\mathbb{R}_+$;

$$\tilde{H} \leq H(t) \leq H(0) = X_0 - \tilde{Y}. \tag{62}$$

Therefore, we can conclude that there exists a constant $C > 0$ such that

$$\|X(t) - \tilde{Y}\|_F \leq Ce^{-4\sigma_{min}t}, \tag{63}$$

where $\sigma_{min} = \min\{|\sigma_i|\}$ is the smallest eigenvalue of Y in absolute value.

C. Rate of Convergence in the Asymmetric Case

Here we prove Proposition 5. First, note that equation $\dot{R} = \tilde{S}R - \frac{1}{2}RR^TR$ follows directly from $\dot{R} = SR - \frac{1}{2}RR^TR + \bar{S}R\Lambda_{\mathcal{Q}_0}$ when $\Lambda_{\mathcal{Q}_0} = \lambda_0 I$. This leads to a Riccati differential equation for $P(t) = R(t)R^T(t)$:

$$\begin{aligned}\frac{dP(t)}{dt} &= \dot{R}R^T + R\dot{R}^T = \left(\tilde{S}R - \tfrac{1}{2}RR^TR\right)R^T + R\left(\tilde{S}R - \tfrac{1}{2}RR^TR\right)^T \\ &= \tilde{S}P(t) + P(t)\tilde{S} - (P(t))^2\end{aligned} \tag{64}$$

with initial conditions $P(0) = R_0R_0^T$. By the arguments used in Appendix A, the exact solution of the above equation is given by;

$$\left(RR^T\right)(t) = e^{t\tilde{S}}R_0R_0^T\left(I + \tfrac{1}{2}\tilde{S}^{-1}(e^{2t\tilde{S}} - I)R_0R_0^T\right)^{-1}e^{t\tilde{S}}. \tag{65}$$

One can show that RR^T converges exponentially to the matrix $R_\star R_\star^T$; defined as the projection of $2\tilde{S}$ on the positive semidefinte cone; this is derived by the same arguments leading to Proposition 2 (see Appendix B). The convergence rate depends on the eigenvalues of $2\tilde{S}$ which can be determined using Schur's complement formula. For a block matrix M one has

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \Longrightarrow \quad \det(M) = \det(D)\det(A - BD^{-1}C). \tag{66}$$

Thus, using the above formula for $\tilde{S} - \lambda I = \left[\begin{smallmatrix} (\lambda_0/2-\lambda)I_m & \Sigma \\ \Sigma^T & -(\lambda_0/2+\lambda+)I_n \end{smallmatrix}\right]$ we have

$$\det(\tilde{S} - \lambda I) = (-1)^{m+n}(\lambda_0/2 + \lambda)^{n-m}\prod_{i=1}^{m}\left(\sigma_i^2 + \lambda_0^2/4 - \lambda^2\right). \tag{67}$$

Therefore, the eigenvalues of $\tilde{S}$ are:

1. $2m$ eigenvalues $\pm\tilde{s}_i$ where $\tilde{s}_i = \frac{1}{2}\sqrt{4\sigma_i^2 + \lambda_0^2}$ $(i = 1, \ldots, m)$.
2. The eigenvalue $-\lambda_0/2$ of multiplicity at least $n - m$.

In general the smallest magnitude eigenvalue will be $|\lambda_0|/2$. However, if Y is a square matrix then only the eigenvalues $\pm\tilde{s}_i$ above will be present.

D. From Implicit Acceleration to Explicit Regularization

We now prove Proposition 4. Replacing (24) into (25) immediately yields (26). It is also straightforward to verify that applying the gradient flow, $\frac{d}{dt}\bar{U} = -\nabla_{\bar{U}}\ell$ and $\frac{d}{dt}\bar{V} = -\nabla_{\bar{V}}\ell$, with ℓ being the objective function in (27), yields (26). Next, consider the formal Taylor series

$$\bar{\mathcal{Q}}(t) = \bar{\mathcal{Q}}(t_0) + (t - t_0)\dot{\bar{\mathcal{Q}}}(t_0) + \tfrac{1}{2}(t - t_0)^2\ddot{\bar{\mathcal{Q}}}(t_0) + \cdots. \tag{68}$$

Define $\mathcal{P}(t) \equiv \bar{U}^T(t)\bar{U}(t) + \bar{V}^T(t)\bar{V}(t)$. From (26) one obtains

$$\frac{d}{dt}\bar{\mathcal{Q}}(t) = -\tfrac{1}{2}(\bar{\mathcal{Q}}(t) - \Lambda_{\mathcal{Q}_0})\mathcal{P}(t) - \tfrac{1}{2}\mathcal{P}(t)(\bar{\mathcal{Q}}(t) - \Lambda_{\mathcal{Q}_0}). \tag{69}$$

Since this is "linear" in $(\bar{\mathcal{Q}} - \Lambda_{\mathcal{Q}_0})$, higher order derivatives take the form

$$\frac{d^n}{dt^n}\bar{\mathcal{Q}}(t) = \sum_i \mathcal{Z}_i(t)\big(\bar{\mathcal{Q}}(t) - \Lambda_{\mathcal{Q}_0}\big)\mathcal{W}_i(t) \tag{70}$$

where the functions $\mathcal{Z}_i$'s and $\mathcal{W}_i$'s contain a sum of powers and time derivatives of $\mathcal{P}(t)$. For instance, the second order derivative yields

$$\begin{aligned}\ddot{\bar{\mathcal{Q}}} = \frac{d^2}{dt^2}\bar{\mathcal{Q}} &= -\frac{1}{2}\dot{\mathcal{Q}}(t)\mathcal{P}(t) - \frac{1}{2}(\mathcal{Q} - \Lambda_{\mathcal{Q}_0})\dot{\mathcal{P}}(t) - \frac{1}{2}\mathcal{P}(t)\dot{\mathcal{Q}}(t) - \frac{1}{2}\dot{\mathcal{P}}(t)(\mathcal{Q} - \Lambda_{\mathcal{Q}_0}) \\ &= -\tfrac{1}{4}\big(\mathcal{Q} - \Lambda_{\mathcal{Q}_0}\big)\big(\mathcal{P}^2 + 2\dot{\mathcal{P}}\big) - \tfrac{1}{4}\big(\mathcal{P}^2 + 2\dot{\mathcal{P}}\big)\big(\mathcal{Q} - \Lambda_{\mathcal{Q}_0}\big) - \tfrac{1}{2}\mathcal{P}\big(\mathcal{Q} - \Lambda_{\mathcal{Q}_0}\big)\mathcal{P}.\end{aligned}$$

Therefore, if $\bar{\mathcal{Q}}(t_0) = \Lambda_{\mathcal{Q}_0}$ at $t = t_0$ then all derivatives (70) vanish identically. As a consequence, the expansion (68) implies $\bar{\mathcal{Q}}(t) = \bar{\mathcal{Q}}(t_0) = \Lambda_{\mathcal{Q}_0}$ for any other $t \geq t_0$ as well.