A. Proof of Theorem 7 (Sampled Positive Definite Quadratics)

We begin by noting that σ is positive definite if and only if the optimal value of the following problem is positive for any value of $\varepsilon > 0$,

$$\inf_{\xi} \ \sigma(\xi) \quad \text{s.t.} \begin{cases} (\mathbf{x}, \mathbf{g}, \mathbf{f}) \text{ as in (16) generated by method } (\mathcal{M}) \\ \quad\quad \text{applied to } f \in \mathcal{F}_{\mu,L}, \\ \|\xi\| \geq \varepsilon. \end{cases}$$

Next, we discretize the problem by replacing $f \in \mathcal{F}_{\mu,L}$ with the equivalent condition that the discrete set of points $\{(y_i, g_i, f_i)\}_{i\in\mathcal{I}_K}$ is $\mathcal{F}_{\mu,L}$-interpolable (recall that $\mathcal{I}_K = \{0, \dots, K, \star\}$). Choosing a specific notion of distance for ξ, we can reformulate the previous statement as verifying that $p_\star^{(d)}(\varepsilon) > 0$ for all $\varepsilon > 0$ where

$$p_\star^{(d)}(\varepsilon) := \min_{\mathbf{x},\mathbf{g},\mathbf{f}} \ \sigma(\xi) \quad \text{s.t.} \begin{cases} \{(y_i, g_i, f_i)\}_{i\in\mathcal{I}_k} \text{ is } \mathcal{F}_{\mu,L}\text{-interpolable}, \\ (\mathbf{x}, \mathbf{g}, \mathbf{f}) \text{ as in (16) generated by } (\mathcal{M}), \\ \|\mathbf{x}\|^2 + \|\mathbf{g}\|^2 + \mathbf{1}^\mathsf{T}\mathbf{f} = \varepsilon. \end{cases}$$

Note that the last condition can equivalently be replaced by others (e.g., $\|\mathbf{x}\|^2 = \varepsilon$ or $\|\mathbf{g}\|^2 = \varepsilon$ or $\mathbf{1}^\mathsf{T}\mathbf{f} = \varepsilon$), and that the optimal value of this problem is attained (using a short homogeneity argument with respect to ε). Using the necessary and sufficient conditions for the set of points $\{(y_i, g_i, f_i)\}_{i\in\mathcal{I}_k}$ to be $\mathcal{F}_{\mu,L}$-interpolable from Theorem 6, we have

$$p_\star^{(d)}(\varepsilon) = \min_{\mathbf{x},\mathbf{g},\mathbf{f}} \ \sigma(\xi) \quad \text{s.t.} \begin{cases} \phi_{ij} \geq 0 \text{ for all } i,j \in \mathcal{I}_K, \\ (\mathbf{x}, \mathbf{g}, \mathbf{f}) \text{ as in (16) generated by } (\mathcal{M}), \\ \|\mathbf{x}\|^2 + \|\mathbf{g}\|^2 + \mathbf{1}^\mathsf{T}\mathbf{f} = \varepsilon. \end{cases}$$

Next, we define the Gram matrix $\mathbf{G} \in \mathbb{S}^{N+K+2}$ as $\mathbf{G} := \mathbf{B}^\mathsf{T}\mathbf{B}$ where

$$\mathbf{B} := \begin{bmatrix} x_{-N} - x_\star & \dots & x_0 - x_\star & g_0 & \dots & g_K \end{bmatrix}, \quad (24)$$

hence $\mathbf{G}$ is a standard Gram matrix containing all inner products between $x_i - x_\star$ for $i \in \{-N, \dots, 0\}$ and g_i for $i \in \{0, \dots, K\}$. Note that the quadratic σ can be written as a function of the Gram matrix as

$$\sigma(\mathbf{G}, \mathbf{f}) = \operatorname{tr}(Q\mathbf{G}) + q^\mathsf{T}\mathbf{f}.$$

Similarly, the interpolation conditions can also be reformulated with the Gram matrix as

$$0 \leq \phi_{ij}(\mathbf{G}, \mathbf{f}) = \operatorname{tr}(M_{ij}\mathbf{G}) + m_{ij}^\mathsf{T}\mathbf{f}$$

where M_{ij} and m_{ij} are such that

$$\phi_{ij} = \begin{bmatrix} \mathbf{x} \\ \mathbf{g} \end{bmatrix}^\mathsf{T} (M_{ij} \otimes I_d) \begin{bmatrix} \mathbf{x} \\ \mathbf{g} \end{bmatrix} + m_{ij}^\mathsf{T}\mathbf{f},$$

for all $i, j \in \mathcal{I}_K$ (hence also $\star$ is in the index set). Therefore, we can reformulate the previous problem as the following rank-constrained semidefinite program:

$$p_\star^{(d)}(\varepsilon) = \min_{\mathbf{G}\in\mathbb{S}^{N+K+2}, \mathbf{f}\in\mathbb{R}^{N+1}} \ \operatorname{tr}(Q\mathbf{G}) + q^\mathsf{T}\mathbf{f} \quad \text{s.t.} \begin{cases} 0 \leq \operatorname{tr}(M_{ij}\mathbf{G}) + m_{ij}^\mathsf{T}\mathbf{f} \quad \text{for } i,j \in \mathcal{I}, \\ \operatorname{tr}(\mathbf{G}) + \mathbf{1}^\mathsf{T}\mathbf{f} = \varepsilon, \\ \mathbf{f} \geq 0, \\ \mathbf{G} \succeq 0, \\ \operatorname{Rank}(\mathbf{G}) \leq d, \end{cases}$$

where we remind the reader that d is the dimension of the optimization problem ($\mathcal{P}$). Therefore, as discussed in (Taylor et al., 2017c;b), if we want a result that does not depend on the dimension (i.e, a σ that is positive definite whatever the value of d), we have to verify that $p^{(\infty)}(\varepsilon) > 0$ (which corresponds to assuming that $d \geq N+K+2$ since this is the dimension of $\mathbf{G}$). We then have the following semidefinite program:

$$p_\star^{(\infty)}(\varepsilon) = \min_{\mathbf{G}\in\mathbb{S}^{N+K+2}, \mathbf{f}\in\mathbb{R}^{N+1}} \ \operatorname{tr}(Q\mathbf{G}) + q^\mathsf{T}\mathbf{f} \quad \text{s.t.} \begin{cases} 0 \leq \operatorname{tr}(M_{ij}\mathbf{G}) + m_{ij}^\mathsf{T}\mathbf{f} \quad \text{for } i,j \in \mathcal{I}, \\ \operatorname{tr}(\mathbf{G}) + \mathbf{1}^\mathsf{T}\mathbf{f} = \varepsilon, \\ \mathbf{f} \geq 0, \\ \mathbf{G} \succeq 0. \end{cases}$$

A Slater point for this problem (i.e., a feasible point such that $\mathbf{G} \succ 0$) is obtained in the following section, so the optimal value of the primal problem is equal to the optimal value of the dual, which is given by

$$d^{(\infty)}(\varepsilon) := \max_{\{\lambda_{ij}\},\nu} \ \nu\,\varepsilon \quad \text{s.t.} \begin{cases} \lambda_{ij} \geq 0 \text{ for all } i,j \in \mathcal{I}, \\ Q - \sum_{i,j\in\mathcal{I}} \lambda_{ij} M_{ij} \succeq \nu I_{N+K+2}, \\ q - \sum_{i,j\in\mathcal{I}} \lambda_{ij} m_{ij} \geq \nu \mathbf{1}_{N+1}. \end{cases}$$

The theorem is then proved by noting the equivalence

$$p^{(\infty)}(\varepsilon) > 0, \ \forall\varepsilon > 0 \iff d^{(\infty)}(\varepsilon) > 0, \ \forall\varepsilon > 0,$$

where the last statement amounts to verifying that

$$Q - \sum_{i,j\in\mathcal{I}} \lambda_{ij} M_{ij} \succ 0, \quad q - \sum_{i,j\in\mathcal{I}} \lambda_{ij} m_{ij} > 0. \quad \blacksquare$$

B. Slater Point for Proof of Theorem 7

In this section, we show how to construct a Slater point (Boyd & Vandenberghe, 2004) for the primal semidefinite program in the proof of Theorem 7. The construction is similar to Section 2.1.2 of (Nesterov, 2004) and the proof of Theorem 6 in (Taylor et al., 2017c).

Consider applying the first-order iterative fixed-step method ($\mathcal{M}$) with $\alpha \neq 0$ and $\gamma_0 \neq 0$ for K iterations to the function $f(x) = \frac{1}{2}x^\top Hx$ where $H \in \mathbb{S}^d$ with $d \geq N + K + 2$ is the positive definite tridiagonal matrix defined by

$$[H]_{ij} = \begin{cases} 2 & \text{if } i = j \\ 1 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

which has maximum eigenvalue $L = 2 + 2\cos\big(\pi/(d+1)\big)$. Define the matrix $\mathbf{B}$ in (24). Using the initial condition $x_i = e_{N+1+i}$ for $i = -N, \dots, 0$, we will show that $\mathbf{B}$ is upper triangular with nonzero diagonal elements, and hence full rank.

Since $\gamma_0 \neq 0$, y_0 has a nonzero element corresponding to e_{N+1}. Then $g_0 = Hy_0$ has a nonzero element corresponding to e_{N+2} due to the tridiagonal structure of H. Furthermore, y_0 may only have nonzero elements corresponding to e_j for $j = 1, \dots, N+1$, so g_0 must have zero components corresponding to e_j for all $j > N+2$.

We now continue by induction. Assume that g_k has a nonzero element corresponding to e_{N+2+k} and zero elements corresponding to e_i for all $i > N+2+k$. Since $\alpha \neq 0$, x_{k+1} has a nonzero element corresponding to e_{N+2+k} and zero elements corresponding to e_i for all $i > N+2+k$. Then since $\gamma_0 \neq 0$, y_{k+1} also has the same structure. Due to the tridiagonal structure of H, g_{k+1} then has a nonzero element corresponding to e_{N+3+k} and zero elements corresponding to e_i for all $i > N+3+k$. Therefore, by induction we have shown that for all $k \geq 0$, g_k has a nonzero element corresponding to e_{N+2+k} and zero elements corresponding to e_i for all $i > N+2+k$. For P to be upper triangular, we need g_K to have dimension at least $N+2+K$. Thus, if $d \geq N+2+K$ where K is the number of iterations, then $\mathbf{B}$ is upper triangular with nonzero entries on the diagonal, and therefore has full rank.

In order to make the statement hold for general $\mu < L$, observe that the tridiagonal structure of H is preserved under the operation

$$\tilde{H} = \big(H - \lambda_{\min}(H)\, I\big)\frac{L-\mu}{\lambda_{\max}(H) - \lambda_{\min}(H)} + \mu I$$

where $\mu I \preceq \tilde{H} \preceq LI$.

Since $\mathbf{B}$ has full rank, the Gram matrix $\mathbf{G} = \mathbf{B}^\top\mathbf{B} \succ 0$ is positive definite. Therefore, the primal semidefinite program satisfies Slater's condition.

C. Steepest Descent

In this section, we show a similar formulation as (ρ-SDP) for steepest descent. In this case, the analysis was not *a priori* guaranteed to be tight, due to the line search conditions. In order to encode the line search, we use the corresponding optimality conditions, as in (de Klerk et al., 2017):

$$\begin{aligned} \langle x_{k+1} - x_k,\ g_{k+1}\rangle &= 0, \\ \langle g_k,\ g_{k+1}\rangle &= 0, \end{aligned} \tag{25}$$

with $g_k = \nabla f(x_k)$. For the Lyapunov function structure, we choose the following

$$V(\xi_k) = \begin{bmatrix} x_k - x_\star \\ g_k \end{bmatrix} (P \otimes I_d) \begin{bmatrix} x_k - x_\star \\ g_k \end{bmatrix}^\top + p(f_k - f_\star).$$

In order to develop a SDP formulation for this problem we follow the same steps as for (ρ-SDP), starting with **Step 1** (see Section 4.2): we define the following row vectors in $\mathbb{R}^2$

$$\bar{y}_0^{(0)} := \mathbf{e}_1^\top,\ \bar{x}_0^{(0)} := \mathbf{e}_1^\top,\ \bar{g}_0^{(0)} := \mathbf{e}_2^\top,$$

and $\bar{y}_\star^{(0)} = \bar{x}_\star^{(0)} = \bar{g}_\star^{(0)} := \mathbf{0}^\top$, along with the scalars $\bar{f}_0^{(0)} := 1$ and $\bar{f}_\star^{(0)} := 0$. In addition, we use the following vectors in $\mathbb{R}^4$

$$\begin{aligned} &\bar{x}_0^{(1)} := \mathbf{e}_1^\top,\ \bar{x}_1^{(1)} := \mathbf{e}_2^\top, \\ &\bar{y}_0^{(1)} := \mathbf{e}_2^\top,\ \bar{y}_1^{(1)} := \mathbf{e}_2^\top, \\ &\bar{g}_0^{(1)} = \mathbf{e}_3^\top,\ \bar{g}_1^{(1)} := \mathbf{e}_4^\top, \end{aligned}$$

and $\bar{y}_\star^{(1)} = \bar{x}_\star^{(1)} = \bar{g}_\star^{(1)} := \mathbf{0}^\top$, along with $\bar{f}_0^{(1)}, \bar{f}_1^{(1)}, \bar{f}_\star^{(1)} \in \mathbb{R}^2$ such that $\bar{f}_0^{(1)} := \mathbf{e}_1^\top$, $\bar{f}_1^{(1)} := \mathbf{e}_2^\top$ and $\bar{f}_\star^{(1)} := \mathbf{0}^\top$.

Because of the algorithm, **Step 2** is slightly different as before; we encode the line search constraints (25) using

$$\begin{aligned} A_1 &= \begin{bmatrix} \bar{x}_0^{(1)} \\ \bar{x}_1^{(1)} \\ \bar{g}_1^{(1)} \end{bmatrix}^\top \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_0^{(1)} \\ \bar{x}_1^{(1)} \\ \bar{g}_1^{(1)} \end{bmatrix}, \\ A_2 &= \begin{bmatrix} \bar{g}_0^{(1)} \\ \bar{g}_1^{(1)} \end{bmatrix}^\top \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{g}_0^{(1)} \\ \bar{g}_1^{(1)} \end{bmatrix}. \end{aligned}$$

The subsequent steps (**Step 3** and **Step 4**) are exactly the same as in Section 4.2. We finally obtain a slightly modified version of the feasibility problem (ρ-SDP):

$$\begin{aligned}
\underset{\substack{P\in\mathbb{S}^2\\ p\in\mathbb{R}^1\\ \nu_1,\nu_2\in\mathbb{R}\\ \{\lambda_{ij}\}\\ \{\eta_{ij}\}}}{\text{feasible}} \quad & 0 \prec V_0^{(0)} - \sum_{i,j\in\mathcal{I}_0} \lambda_{ij}\, M_{ij}^{(0)} \\
& 0 < v_0^{(0)} - \sum_{i,j\in\mathcal{I}_0} \lambda_{ij}\, m_{ij}^{(0)} \\
& 0 \succeq \Delta V_0^{(1)} + \sum_{i,j\in\mathcal{I}_1} \eta_{ij}\, M_{ij}^{(1)} + \sum_{i=1}^{2} \nu_i\, A_i \\
& 0 \geq \Delta v_0^{(1)} + \sum_{i,j\in\mathcal{I}_1} \eta_{ij}\, m_{ij}^{(1)} \\
& 0 \leq \lambda_{ij} \quad \text{for } i,j\in\mathcal{I}_0 \\
& 0 \leq \eta_{ij} \quad \text{for } i,j\in\mathcal{I}_1
\end{aligned}$$

with $\mathcal{I}_0 := \{0,\star\}$ and $\mathcal{I}_1 := \{0,1,\star\}$.

D. SDP for HBM with Subspace Searches

We follow the steps of the previous section for steepest descent; we only make the following adaptations: (i) we look for a quadratic Lyapunov function with the states

$$V(\xi_k) = \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ g_k \\ g_{k-1} \end{bmatrix} (P\otimes I_d) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ g_k \\ g_{k-1} \end{bmatrix}^\top + p^\top \begin{bmatrix} f_k - f_\star \\ f_{k-1} - f_\star \end{bmatrix},$$

(ii) we adapt the initialization (**Step 1**), (iii) adapt the line search conditions (**Step 2**) and (iv) obtain a slightly modified version of the SDP.

For (ii), we adapt the initialization procedure (**Step 1**) as follows. We define the following row vectors of $\mathbb{R}^4$:

$$\begin{aligned}
\bar{x}_0^{(1)} &:= \mathbf{e}_1^\top,\ \bar{x}_1^{(1)} := \mathbf{e}_2^\top, \\
\bar{y}_0^{(1)} &:= \mathbf{e}_1^\top,\ \bar{y}_1^{(1)} := \mathbf{e}_2^\top, \\
\bar{g}_0^{(1)} &:= \mathbf{e}_3^\top,\ \bar{g}_1^{(1)} := \mathbf{e}_4^\top,
\end{aligned}$$

along with $\bar{y}_\star^{(1)} = \bar{x}_\star^{(1)} = \bar{g}_\star^{(1)} := \mathbf{0}^\top$, and the following in $\mathbb{R}^2$: $\bar{f}_0^{(1)} := \mathbf{e}_1^\top$, $\bar{f}_1^{(1)} := \mathbf{e}_2^\top$ and $\bar{f}_\star^{(1)} := \mathbf{0}^\top$. We also define the following row vectors in $\mathbb{R}^7$:

$$\begin{aligned}
\bar{x}_{-1}^{(2)} := \mathbf{e}_1^\top,\ \bar{x}_0^{(2)} := \mathbf{e}_2^\top,\ \bar{x}_1^{(2)} := \mathbf{e}_3^\top,\ \bar{x}_2^{(2)} := \mathbf{e}_4^\top, \\
\bar{y}_0^{(2)} := \bar{x}_0^{(2)},\ \bar{y}_1^{(2)} := \bar{x}_1^{(2)},\ \bar{y}_2^{(2)} := \bar{x}_2^{(2)}, \\
\bar{g}_0^{(2)} := \mathbf{e}_5^\top,\ \bar{g}_1^{(2)} := \mathbf{e}_6^\top,\ \bar{g}_2^{(2)} := \mathbf{e}_7^\top,
\end{aligned}$$

along with $y_\star^{(2)} = \bar{x}_\star^{(2)} = \bar{g}_\star^{(2)} := \mathbf{0}^\top$ and the vectors of $\mathbb{R}^3$: $\bar{f}_0^{(2)} := \mathbf{e}_1^\top$, $\bar{f}_1^{(2)} := \mathbf{e}_2^\top$, $\bar{f}_2^{(2)} := \mathbf{e}_3^\top$ and $\bar{f}_\star^{(2)} := \mathbf{0}^\top$.

Now for (iii) (or **Step 2**), optimality of the search conditions can be

$$\begin{aligned}
\langle x_{k+1} - x_k,\ g_{k+1}\rangle &= 0, \\
\langle x_k - x_{k-1},\ g_{k+1}\rangle &= 0, \\
\langle g_k; g_{k+1}\rangle &= 0,
\end{aligned}$$

which we can formulate in matrix form for $k\in\{0,1\}$:

$$\begin{aligned}
A_{1+k} &= \begin{bmatrix} \bar{x}_k^{(2)} \\ \bar{x}_{k+1}^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix}^\top \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_k^{(2)} \\ \bar{x}_{k+1}^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix} \\
A_{3+k} &= \begin{bmatrix} \bar{x}_{k-1}^{(2)} \\ \bar{x}_{k}^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix}^\top \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_{k-1}^{(2)} \\ \bar{x}_{k}^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix} \\
A_{5+k} &= \begin{bmatrix} \bar{g}_k^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix}^\top \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{g}_k^{(2)} \\ \bar{g}_{k+1}^{(2)} \end{bmatrix}
\end{aligned}$$

The subsequent steps (**Step 3** and **Step 4**) are exactly the same as in Section 4.2. We finally obtain a slightly modified version of the feasibility problem (ρ-SDP):

$$\begin{aligned}
\underset{\substack{P\in\mathbb{S}^2\\ p\in\mathbb{R}^1\\ \nu_1,\dots,\nu_6\in\mathbb{R}\\ \{\lambda_{ij}\}\\ \{\eta_{ij}\}}}{\text{feasible}} \quad & 0 \prec V_1^{(1)} - \sum_{i,j\in\mathcal{I}_1} \lambda_{ij}\, M_{ij}^{(1)} \\
& 0 < v_1^{(1)} - \sum_{i,j\in\mathcal{I}_1} \lambda_{ij}\, m_{ij}^{(1)} \\
& 0 \succeq \Delta V_1^{(2)} + \sum_{i,j\in\mathcal{I}_2} \eta_{ij}\, M_{ij}^{(2)} + \sum_{i=1}^{6} \nu_i\, A_i \\
& 0 \geq \Delta v_1^{(2)} + \sum_{i,j\in\mathcal{I}_2} \eta_{ij}\, m_{ij}^{(2)} \\
& 0 \leq \lambda_{ij} \quad \text{for } i,j\in\mathcal{I}_1 \\
& 0 \leq \eta_{ij} \quad \text{for } i,j\in\mathcal{I}_2
\end{aligned}$$

with $\mathcal{I}_1 := \{0,1,\star\}$ and $\mathcal{I}_2 := \{0,1,2,\star\}$. The corresponding results are presented on Figure 3.

E. SDP for FGM with Scheduled Restarts

This setting goes slightly beyond the fixed-step model presented in ($\mathcal{M}$), as the step sizes depend on the iteration. We study the algorithm described by (23), which does N steps of the standard fast gradient method (Nesterov, 1983) before restarting. We study the convergence of this scheme using quadratic Lyapunov functions of the form

$$\begin{bmatrix} y_k^N - x_\star \\ \nabla f(y_k^N) \end{bmatrix}^\top (P\otimes I_d) \begin{bmatrix} y_k^N - x_\star \\ \nabla f(y_k^N) \end{bmatrix} + p\,[f(y_k^N) - f(x_\star)]. \tag{26}$$

Let us perform similar steps as for (ρ-SDP) for constructing the corresponding SDP. We start with the initialization procedure **Step 1**. Let us define the following row vectors in $\mathbb{R}^2$:

$$\bar{y}_0^{(1)} = \bar{x}_0^{(1)} := \mathbf{e}_1^\top,\, \bar{g}_0^{(1)} := \mathbf{e}_2^\top,$$

and $\bar{x}_\star^{(1)} = \bar{g}_\star^{(1)} := \mathbf{0}^\top$, along with the scalars $f_0^{(1)} := 1$ and $f_\star^{(1)} := 0$. In addition, we define the row vectors of $\mathbb{R}^{N+2}$:

$$\bar{x}_0^{(N+1)} := \mathbf{e}_1^\top,\, \bar{g}_k^{(N+1)} := \mathbf{e}_{2+k}^\top,$$

for $k = 0, \ldots, N$, along with $\bar{x}_\star^{(N+1)} = \bar{g}_\star^{(N+1)} := \mathbf{0}^\top$ and the row vectors $\bar{f}_k^{(N+1)} \in \mathbb{R}^{N+1}$ defined as $\bar{f}_k^{(N+1)} := \mathbf{e}_{1+k}^\top$ and $\bar{f}_\star^{(N+1)} := \mathbf{0}^\top$.

Step 2 Apply one complete loop of the algorithm as follows: for $k = 0, \ldots, N-1$ define the sequence of row vectors:

$$\begin{aligned}
\bar{z}_{k+1}^{(N+1)} &= \bar{y}_k^{(N+1)} - \frac{1}{L}\bar{g}_k^{(N+1)},\\
\bar{y}_{k+1}^{(N+1)} &= \bar{z}_k^{(N+1)} + \frac{\beta_k - 1}{\beta_{k+1}}(\bar{z}_k^{(N+1)} - \bar{z}_k^{(N+1)}),
\end{aligned}$$

with $\beta_0 := 1$ and $\beta_{k+1} := \frac{1+\sqrt{4\beta_k^2+1}}{2}$. For complying with the notations of the paper, we define the sequence

$$x_k^{(K)} := y_k^{(K)}$$

for $k = 0, \ldots, N$ and $K \in \{1, N+1\}$. Then, using the sets $\mathcal{I}_1 := \{0, \star\}$ and $\mathcal{I}_{N+1} := \{0, \ldots, N, \star\}$, the other stages follow from the same lines as **Step 3**, and **Step 4** with the slight modification of the expression for the rate

$$\begin{aligned}
\Delta v_k^{(N+1)} &:= v_{k+1}^{(N+1)} - \rho^{2N}\, v_k^{(N+1)},\\
\Delta V_k^{(N+1)} &:= V_{k+1}^{(N+1)} - \rho^{2N}\, V_k^{(N+1)},
\end{aligned}$$

and **Step 5** follows as in Section 4.2:

$$\begin{aligned}
\underset{\substack{P\in\mathbb{S}^{2(N+1)}\\ p\in\mathbb{R}^{N+1}\\ \{\lambda_{ij}\}\\ \{\eta_{ij}\}}}{\text{feasible}} \quad & 0 \prec V_1^{(1)} - \sum_{i,j\in\mathcal{I}_1} \lambda_{ij}\, M_{ij}^{(1)}\\
& 0 < v_1^{(1)} - \sum_{i,j\in\mathcal{I}_1} \lambda_{ij}\, m_{ij}^{(1)}\\
& 0 \succeq \Delta V_N^{(N+1)} + \sum_{i,j\in\mathcal{I}_{N+1}} \eta_{ij}\, M_{ij}^{(N+1)}\\
& 0 \geq \Delta v_N^{(N+1)} + \sum_{i,j\in\mathcal{I}_{N+1}} \eta_{ij}\, m_{ij}^{(N+1)}\\
& 0 \leq \lambda_{ij} \quad \text{for } i,j\in\mathcal{I}_1\\
& 0 \leq \eta_{ij} \quad \text{for } i,j\in\mathcal{I}_{N+1}
\end{aligned}$$

(note that the sets $\mathcal{I}_1$ and $\mathcal{I}_{N+1}$ should use the definitions of this section). Numerical results are available in Figure 4.