Details for Deriving the Optimal Convergence Rates

Sandra S. Y. Tan

This document is intended to be supplementary material for the paper "A Unified Framework for the Convergence Analysis of Optimization Algorithms via Sum-of-Squares". It provides full details for the proving the results given in that paper.

1 Full Details for PGM with Constant Step Size

The constraints characterizing the function class and algorithm are

$$
\begin{aligned}
g_1: &\quad \mathbf{x}_{k+1} - \mathbf{x}_k + \gamma(\mathbf{g}_k + \mathbf{s}_{k+1}) = 0\\
g_2: &\quad \mathbf{g}_* + \mathbf{s}_* = 0\\
h_1: &\quad f_k - f_{k+1} - \mathbf{g}_{k+1}^T(\mathbf{x}_k - \mathbf{x}_{k+1}) - \alpha\left[\frac{1}{L}\left\|\mathbf{g}_k - \mathbf{g}_{k+1}\right\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_{k+1} - \mathbf{g}_k)^T(\mathbf{x}_{k+1} - \mathbf{x}_k)\right] \geq 0\\
h_2: &\quad f_k - f_* - \mathbf{g}_*^T(\mathbf{x}_k - \mathbf{x}_*) - \alpha\left[\frac{1}{L}\|\mathbf{g}_k - \mathbf{g}_*\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_*\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_* - \mathbf{g}_k)^T(\mathbf{x}_* - \mathbf{x}_k)\right] \geq 0\\
h_3: &\quad f_{k+1} - f_k - \mathbf{g}_k^T(\mathbf{x}_{k+1} - \mathbf{x}_k) - \alpha\left[\frac{1}{L}\left\|\mathbf{g}_{k+1} - \mathbf{g}_k\right\|^2 + \mu\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_k - \mathbf{g}_{k+1})^T(\mathbf{x}_k - \mathbf{x}_{k+1})\right] \geq 0\\
h_4: &\quad f_{k+1} - f_* - \mathbf{g}_*^T(\mathbf{x}_{k+1} - \mathbf{x}_*) - \alpha\left[\frac{1}{L}\left\|\mathbf{g}_{k+1} - \mathbf{g}_*\right\|^2 + \mu\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_* - \mathbf{g}_{k+1})^T(\mathbf{x}_* - \mathbf{x}_{k+1})\right] \geq 0\\
h_5: &\quad f_* - f_k - \mathbf{g}_k^T(\mathbf{x}_* - \mathbf{x}_k) - \alpha\left[\frac{1}{L}\|\mathbf{g}_* - \mathbf{g}_k\|^2 + \mu\|\mathbf{x}_* - \mathbf{x}_k\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_k - \mathbf{g}_*)^T(\mathbf{x}_k - \mathbf{x}_*)\right] \geq 0\\
h_6: &\quad f_* - f_{k+1} - \mathbf{g}_{k+1}^T(\mathbf{x}_* - \mathbf{x}_{k+1}) - \alpha\left[\frac{1}{L}\left\|\mathbf{g}_* - \mathbf{g}_{k+1}\right\|^2 + \mu\|\mathbf{x}_* - \mathbf{x}_{k+1}\|^2\right.\\
&\quad \left. - 2\frac{\mu}{L}(\mathbf{g}_{k+1} - \mathbf{g}_*)^T(\mathbf{x}_{k+1} - \mathbf{x}_*)\right] \geq 0\\
h_7: &\quad h_k - h_{k+1} - \mathbf{s}_{k+1}^T(\mathbf{x}_k - \mathbf{x}_{k+1}) \geq 0\\
h_8: &\quad h_k - h_* - \mathbf{s}_*^T(\mathbf{x}_k - \mathbf{x}_*) \geq 0\\
h_9: &\quad h_{k+1} - h_k - \mathbf{s}_k^T(\mathbf{x}_{k+1} - \mathbf{x}_k) \geq 0\\
h_{10}: &\quad h_{k+1} - h_* - \mathbf{s}_*^T(\mathbf{x}_{k+1} - \mathbf{x}_*) \geq 0\\
h_{11}: &\quad h_* - h_k - \mathbf{s}_k^T(\mathbf{x}_* - \mathbf{x}_k) \geq 0\\
h_{12}: &\quad h_* - h_{k+1} - \mathbf{s}_{k+1}^T(\mathbf{x}_* - \mathbf{x}_{k+1}) \geq 0
\end{aligned}
$$

where $\gamma = \frac{2}{L+\mu}$. Substituting g_1 and g_2 into the other constraints, we eliminate variables $\mathbf{x}_{k+1}$ and $\mathbf{s}_*$. Taking $n=1$, we match coefficients of the monomials in

$$\begin{aligned} p(\mathbf{z}) &:= t(f_k + h_k) + (1-t)(f_* + h_*) - (f_{k+1} + h_{k+1}) \\ &= \mathbf{z}_1^T Q \mathbf{z}_1 + \sum_{i=1}^{12} \lambda_i h_i(\mathbf{z}) \end{aligned}$$

to obtain the following set of linear equalities

$$\begin{aligned}
& & 0 &= Q(1,1) \\
&f_*: & 1-t &= 2Q(1,2) - \lambda_2 - \lambda_4 + \lambda_5 + \lambda_6 \\
&f_k: & t &= 2Q(1,3) + \lambda_1 + \lambda_2 - \lambda_3 - \lambda_5 \\
&f_{k+1}: & -1 &= 2Q(1,4) - \lambda_1 + \lambda_3 + \lambda_4 - \lambda_6 \\
&h_*: & 1-t &= 2Q(1,5) - \lambda_8 - \lambda_{10} + \lambda_{11} + \lambda_{12} \\
&h_k: & t &= 2Q(1,6) + \lambda_7 + \lambda_8 - \lambda_9 - \lambda_{11} \\
&h_{k+1}: & -1 &= 2Q(1,7) - \lambda_7 + \lambda_9 + \lambda_{10} - \lambda_{12} \\
& & 0 &= Q(1,i) \quad i = 8,\dots,14 \\
& & 0 &= Q(i,j) = Q(j,i) \quad i = 2,\dots,7,\ j = i,\dots,14 \\
&x_*^2: & 0 &= Q(8,8) - \alpha\mu\,(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&x_* x_k: & 0 &= 2Q(8,9) + 2\alpha\mu\,(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&x_* g_*: & 0 &= 2Q(8,10) + \lambda_2 + \lambda_4 - \lambda_8 - \lambda_{10} + 2\alpha\frac{\mu}{L}(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&x_* g_k: & 0 &= 2Q(8,11) - 2\alpha\mu\left[\frac{1}{L}(\lambda_2 + \lambda_5) + \gamma(\lambda_4 + \lambda_6)\right] - \lambda_5 \\
&x_* g_{k+1}: & 0 &= 2Q(8,12) - 2\alpha\frac{\mu}{L}(\lambda_4 + \lambda_6) - \lambda_6 \\
&x_* s_k: & 0 &= 2Q(8,13) - \lambda_{11} \\
&x_* s_{k+1}: & 0 &= 2Q(8,14) - 2\alpha\mu\gamma(\lambda_4 + \lambda_6) - \lambda_{12} \\
&x_k^2: & 0 &= Q(9,9) - \alpha\mu(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&x_k g_*: & 0 &= 2Q(9,10) - \lambda_2 - \lambda_4 + \lambda_8 + \lambda_{10} - 2\alpha\frac{\mu}{L}(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&x_k g_k: & 0 &= 2Q(9,11) + 2\alpha\mu\left[\frac{1}{L}(\lambda_2 + \lambda_5) + \gamma(\lambda_4 + \lambda_6)\right] + \lambda_5 \\
&x_k g_{k+1}: & 0 &= 2Q(9,12) + 2\alpha\frac{\mu}{L}(\lambda_4 + \lambda_6) + \lambda_6 \\
&x_k s_k: & 0 &= 2Q(9,13) + \lambda_{11} \\
&x_k s_{k+1}: & 0 &= 2Q(9,14) + 2\alpha\mu\gamma(\lambda_4 + \lambda_6) + \lambda_{12} \\
&g_*^2: & 0 &= Q(10,10) - \frac{\alpha}{L}(\lambda_2 + \lambda_4 + \lambda_5 + \lambda_6) \\
&g_* g_k: & 0 &= 2Q(10,11) + 2\frac{\alpha}{L}\left[\lambda_2 + \lambda_5 + \mu\gamma(\lambda_4 + \lambda_6)\right] + \gamma(\lambda_4 - \lambda_{10}) \\
&g_* g_{k+1}: & 0 &= 2Q(10,12) + 2\frac{\alpha}{L}(\lambda_4 + \lambda_6) \\
&g_* s_k: & 0 &= Q(10,13) \\
&g_* s_{k+1}: & 0 &= 2Q(10,14) + \gamma(\lambda_4 - \lambda_{10}) + 2\alpha\frac{\mu}{L}\gamma(\lambda_4 + \lambda_6) \\
&g_k^2: & 0 &= 2Q(11,11) - \frac{\alpha}{L}(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_5) \\
& & &\quad - \alpha\mu\gamma^2(\lambda_1 + \lambda_3 + \lambda_4 + \lambda_6) + 2\alpha\gamma\frac{\mu}{L}(\lambda_1 + \lambda_3) + \gamma\lambda_3
\end{aligned} \tag{1}$$

$$
\begin{aligned}
g_k g_{k+1}: &\quad 0 = 2Q(11,12) - \gamma(\lambda_1+\lambda_6) + 2\frac{\alpha}{L}\left[\lambda_1+\lambda_3-\mu\gamma(\lambda_1+\lambda_3+\lambda_4+\lambda_6)\right] \\
g_k s_k: &\quad 0 = 2Q(11,13) + \gamma\lambda_9 \\
g_k s_{k+1}: &\quad 0 = 2Q(11,14) + 2\alpha\gamma\frac{\mu}{L}(\lambda_1+\lambda_3) + \gamma(\lambda_3-\lambda_7-\lambda_{12}) \\
&\qquad - 2\alpha\mu\gamma^2(\lambda_1+\lambda_3+\lambda_4+\lambda_6) \\
g_{k+1}^2: &\quad 0 = Q(12,12) - \frac{\alpha}{L}(\lambda_1+\lambda_3+\lambda_4+\lambda_6) \\
g_{k+1}s_k: &\quad 0 = Q(12,13) \\
g_{k+1}s_{k+1}: &\quad 0 = 2Q(12,14) - \gamma(\lambda_1+\lambda_6) - 2\alpha\gamma\frac{\mu}{L}(\lambda_1+\lambda_3+\lambda_4+\lambda_6) \\
s_k^2: &\quad 0 = Q(13,13) \\
s_k s_{k+1}: &\quad 0 = 2Q(13,14) + \gamma\lambda_9 \\
s_{k+1}^2: &\quad 0 = Q(14,14) - \alpha\mu\gamma^2(\lambda_1+\lambda_3+\lambda_4+\lambda_6) - \gamma(\lambda_7+\lambda_{12})
\end{aligned}
$$

The SDP is thus minimizing t subject to these linear equalities as well as

$$
\begin{aligned}
&Q \succeq 0 \\
&\lambda_i \geq 0 \quad i = 0,\dots,12 \\
&0 \leq t \leq 1
\end{aligned}
$$

For $\gamma = \frac{2}{L+\mu}$, the optimal values of t and λ's correspond to

$$
t = \left(\frac{L-\mu}{L+\mu}\right)^2 \qquad \lambda_1 = \frac{L-\mu}{L+\mu} \qquad \lambda_5 = 2\mu\frac{L-\mu}{(L+\mu)^2}
$$

$$
\lambda_6 = \frac{2\mu}{L+\mu} \qquad \lambda_7 = \left(\frac{L-\mu}{L+\mu}\right)^2 \qquad \lambda_{12} = \frac{4L\mu}{(L+\mu)^2}
$$

and $\lambda_i = 0$ for $i = 2,\dots,4,8,\dots,11$. The optimal Q is a sparse symmetric matrix, with the non-zero entries in the upper triangular portion being:

$$
Q(8,8) = \frac{2L^2\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(8,9) = -\frac{2L^2\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(8,10) = -\frac{2L\mu^2}{(L+\mu)^2(L-\mu)}
$$

$$
Q(8,11) = \frac{L\mu}{(L+\mu)(L-\mu)} \qquad Q(8,12) = \frac{L\mu}{(L+\mu)(L-\mu)} \qquad Q(8,14) = \frac{2L^2\mu}{(L+\mu)^2(L-\mu)}
$$

$$
Q(9,9) = \frac{2L^2\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(9,10) = \frac{2L\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(9,11) = -\frac{L\mu}{(L+\mu)(L-\mu)}
$$

$$
Q(9,12) = -\frac{L\mu}{(L+\mu)(L-\mu)} \qquad Q(9,14) = -\frac{2L^2\mu}{(L+\mu)^2(L-\mu)} \qquad Q(10,10) = \frac{2L\mu}{(L+\mu)^2(L-\mu)}
$$

$$
Q(10,11) = -\frac{\mu}{(L+\mu)(L-\mu)} \qquad Q(10,12) = -\frac{\mu}{(L+\mu)(L-\mu)} \qquad Q(10,14) = -\frac{2\mu^2}{(L+\mu)^2(L-\mu)}
$$

$$
Q(11,11) = \frac{1}{2(L-\mu)} \qquad Q(11,12) = \frac{1}{2(L-\mu)} \qquad Q(11,14) = \frac{L}{(L+\mu)(L-\mu)}
$$

$$
Q(12,12) = \frac{1}{2(L-\mu)} \qquad Q(12,14) = \frac{L}{(L+\mu)(L-\mu)} \qquad Q(14,14) = 2\frac{L^2+L\mu-\mu^2}{(L+\mu)^2(L-\mu)}
$$

These fulfill the set of linear equalities in the SDP. Q is also PSD, as can be verified by hand or by symbolic computation. This completes the proof that

$$
f_{k+1} - f_* \leq \left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*)
$$

2 Full Details for PGM with Exact Line Search

The constraints characterizing the function class and algorithm are

$$
\begin{aligned}
g_1: \quad & \mathbf{g}_{k+1}^T\mathbf{g}_k + \mathbf{g}_{k+1}^T\mathbf{s}_{k+1} + \mathbf{s}_{k+1}^T\mathbf{g}_k + \|\mathbf{s}_{k+1}\|^2 = 0\\
g_2: \quad & \mathbf{g}_{k+1}^T\mathbf{x}_{k+1} - \mathbf{g}_{k+1}^T\mathbf{x}_k + \mathbf{s}_{k+1}^T\mathbf{x}_{k+1} - \mathbf{s}_{k+1}^T\mathbf{x}_k = 0\\
g_3: \quad & \mathbf{g}_* + \mathbf{s}_* = 0\\
h_4: \quad & f_k - f_{k+1} - \mathbf{g}_{k+1}^T(\mathbf{x}_k - \mathbf{x}_{k+1}) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_k - \mathbf{g}_{k+1}\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_{k+1} - \mathbf{g}_k)^T(\mathbf{x}_{k+1} - \mathbf{x}_k)\Big] \geq 0\\
h_5: \quad & f_k - f_* - \mathbf{g}_*^T(\mathbf{x}_k - \mathbf{x}_*) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_k - \mathbf{g}_*\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_*\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_* - \mathbf{g}_k)^T(\mathbf{x}_* - \mathbf{x}_k)\Big] \geq 0\\
h_6: \quad & f_{k+1} - f_k - \mathbf{g}_k^T(\mathbf{x}_{k+1} - \mathbf{x}_k) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_{k+1} - \mathbf{g}_k\|^2 + \mu\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_k - \mathbf{g}_{k+1})^T(\mathbf{x}_k - \mathbf{x}_{k+1})\Big] \geq 0\\
h_7: \quad & f_{k+1} - f_* - \mathbf{g}_*^T(\mathbf{x}_{k+1} - \mathbf{x}_*) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_{k+1} - \mathbf{g}_*\|^2 + \mu\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_* - \mathbf{g}_{k+1})^T(\mathbf{x}_* - \mathbf{x}_{k+1})\Big] \geq 0\\
h_8: \quad & f_* - f_k - \mathbf{g}_k^T(\mathbf{x}_* - \mathbf{x}_k) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_* - \mathbf{g}_k\|^2 + \mu\|\mathbf{x}_* - \mathbf{x}_k\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_k - \mathbf{g}_*)^T(\mathbf{x}_k - \mathbf{x}_*)\Big] \geq 0\\
h_9: \quad & f_* - f_{k+1} - \mathbf{g}_{k+1}^T(\mathbf{x}_* - \mathbf{x}_{k+1}) - \alpha\Big[\frac{1}{L}\|\mathbf{g}_* - \mathbf{g}_{k+1}\|^2 + \mu\|\mathbf{x}_* - \mathbf{x}_{k+1}\|^2\\
& - 2\frac{\mu}{L}(\mathbf{g}_{k+1} - \mathbf{g}_*)^T(\mathbf{x}_{k+1} - \mathbf{x}_*)\Big] \geq 0\\
h_{10}: \quad & h_k - h_{k+1} - \mathbf{s}_{k+1}^T(\mathbf{x}_k - \mathbf{x}_{k+1}) \geq 0\\
h_{11}: \quad & h_k - h_* - \mathbf{s}_*^T(\mathbf{x}_k - \mathbf{x}_*) \geq 0\\
h_{12}: \quad & h_{k+1} - h_k - \mathbf{s}_k^T(\mathbf{x}_{k+1} - \mathbf{x}_k) \geq 0\\
h_{13}: \quad & h_{k+1} - h_* - \mathbf{s}_*^T(\mathbf{x}_{k+1} - \mathbf{x}_*) \geq 0\\
h_{14}: \quad & h_* - h_k - \mathbf{s}_k^T(\mathbf{x}_* - \mathbf{x}_k) \geq 0\\
h_{15}: \quad & h_* - h_{k+1} - \mathbf{s}_{k+1}^T(\mathbf{x}_* - \mathbf{x}_{k+1}) \geq 0
\end{aligned}
$$

Again, we take $n = 1$ and match coefficients of the monomials in

$$
\begin{aligned}
p(\mathbf{z}) &:= t(f_k + h_k) + (1-t)(f_* + h_*) - (f_{k+1} + h_{k+1})\\
&= \mathbf{z}_1^T Q \mathbf{z}_1 + \sum_{j=1}^{3} \lambda_j g_j(\mathbf{z}) + \sum_{i=4}^{15} \lambda_i h_i(\mathbf{z})
\end{aligned}
$$

to obtain a set of linear equalities. The resulting SDP thus seeks to minimize t subject to

$$
\begin{aligned}
& 0 \leq t \leq 1\\
& \lambda_i \geq 0 \quad i = 4, \ldots, 16\\
& Q \succeq 0\\
& 0 = Q(1,1)
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
&f_* : && 1-t = 2Q(1,2) - \lambda_5 - \lambda_7 + \lambda_8 + \lambda_9 \\
&f_k : && t = 2Q(1,3) + \lambda_4 + \lambda_5 - \lambda_6 - \lambda_8 \\
&f_{k+1} : && -1 = 2Q(1,4) - \lambda_4 + \lambda_6 + \lambda_7 - \lambda_9 \\
&h_* : && 1-t = 2Q(1,5) - \lambda_{11} - \lambda_{13} + \lambda_{14} + \lambda_{15} \\
&h_k : && t = 2Q(1,6) + \lambda_{10} + \lambda_{11} - \lambda_{12} - \lambda_{14} \\
&h_{k+1} : && -1 = 2Q(1,7) - \lambda_{10} + \lambda_{12} + \lambda_{13} - \lambda_{15} \\
& && 0 = Q(1,i) \quad i = 8,9,10,12,13,15,16 \\
&g_* : && 0 = 2Q(1,11) + \lambda_3 \\
&s_* : && 0 = 2Q(1,14) + \lambda_3 \\
& && 0 = Q(i,j) \quad i = 2,\dots,7,\ j = i,\dots,16 \\
&x_*^2 : && 0 = Q(8,8) - \alpha\mu\,(\lambda_5 + \lambda_7 + \lambda_8 + \lambda_9) \\
&x_* x_k : && 0 = 2Q(8,9) + 2\alpha\mu\,(\lambda_5 + \lambda_8) \\
&x_* x_{k+1} : && 0 = 2Q(8,10) + 2\alpha\mu(\lambda_7 + \lambda_9) \\
&x_* g_* : && 0 = 2Q(8,11) + \lambda_5 + \lambda_7 + 2\alpha\frac{\mu}{L}(\lambda_5 + \lambda_7 + \lambda_8 + \lambda_9) \\
&x_* g_k : && 0 = 2Q(8,12) - 2\alpha\frac{\mu}{L}(\lambda_5 + \lambda_8) - \lambda_8 \\
&x_* g_{k+1} : && 0 = 2Q(8,13) - 2\alpha\frac{\mu}{L}(\lambda_7 + \lambda_9) - \lambda_9 \\
&x_* s_* : && 0 = 2Q(8,14) + \lambda_{11} + \lambda_{13} \\
&x_* s_k : && 0 = 2Q(8,15) - \lambda_{14} \\
&x_* s_{k+1} : && 0 = 2Q(8,16) - \lambda_{15} \\
&x_k^2 : && 0 = Q(9,9) - \alpha\mu(\lambda_4 + \lambda_5 + \lambda_6 + \lambda_8) \\
&x_k x_{k+1} : && 0 = 2Q(9,10) + 2\alpha\mu(\lambda_4 + \lambda_6) \\
&x_k g_* : && 0 = 2Q(9,11) - \lambda_5 - 2\alpha\frac{\mu}{L}(\lambda_5 - \lambda_8) \\
&x_k g_k : && 0 = 2Q(9,12) + 2\alpha\frac{\mu}{L}(\lambda_4 + \lambda_5 + \lambda_6 + \lambda_8) + \lambda_6 + \lambda_8 \\
&x_k g_{k+1} : && 0 = 2Q(9,13) - \lambda_2 - \lambda_4 - 2\alpha\frac{\mu}{L}(\lambda_4 + \lambda_6) \\
&x_k s_* : && 0 = 2Q(9,14) - \lambda_{11} \\
&x_k s_k : && 0 = 2Q(9,15) + \lambda_{12} + \lambda_{14} \\
&x_k s_{k+1} : && 0 = 2Q(9,16) - \lambda_2 - \lambda_{10} \\
&x_{k+1}^2 : && 0 = Q(10,10) - \alpha\mu(\lambda_4 + \lambda_6 + \lambda_7 + \lambda_9) \\
&x_{k+1} g_* : && 0 = 2Q(10,11) - \lambda_7 - 2\alpha\frac{\mu}{L}(\lambda_7 + \lambda_9) \\
&x_{k+1} g_k : && 0 = 2Q(10,12) - 2\alpha\frac{\mu}{L}(\lambda_4 + \lambda_6) - \lambda_6 \\
&x_{k+1} g_{k+1} : && 0 = 2Q(10,13) + \lambda_2 + \lambda_4 + \lambda_9 + 2\alpha\frac{\mu}{L}\,(\lambda_4 + \lambda_6 + \lambda_7 + \lambda_9) \\
&x_{k+1} s_* : && 0 = 2Q(10,14) - \lambda_{13} \\
&x_{k+1} s_k : && 0 = 2Q(10,15) - \lambda_{12} \\
&x_{k+1} s_{k+1} : && 0 = 2Q(10,16) + \lambda_2 + \lambda_{10} + \lambda_{15} \\
&g_*^2 : && 0 = Q(11,11) - \frac{\alpha}{L}(\lambda_5 + \lambda_7 + \lambda_8 + \lambda_9) \\
&g_* g_k : && 0 = 2Q(11,12) + 2\frac{\alpha}{L}(\lambda_5 + \lambda_8) \\
&g_* g_{k+1} : && 0 = 2Q(11,13) + 2\frac{\alpha}{L}\,(\lambda_7 + \lambda_9)
\end{aligned}
$$

$$
\begin{aligned}
& & 0 &= Q(11,i) \quad i = 14,\dots,16 \\
& g_k^2: & 0 &= Q(12,12) - \frac{\alpha}{L}(\lambda_4+\lambda_5+\lambda_6+\lambda_8) \\
& g_k g_{k+1}: & 0 &= 2Q(12,13) + \lambda_1 + 2\frac{\alpha}{L}(\lambda_4+\lambda_6) \\
& & 0 &= Q(12,i) \quad i = 14,15 \\
& g_k s_{k+1}: & 0 &= 2Q(12,16) + \lambda_1 \\
& g_{k+1}^2: & 0 &= Q(13,13) - \frac{\alpha}{L}(\lambda_4+\lambda_6+\lambda_7+\lambda_9) \\
& & 0 &= Q(13,i) \quad i = 14,15 \\
& g_{k+1} s_{k+1}: & 0 &= 2Q(13,16) + \lambda_1 \\
& & 0 &= Q(i,j) \quad i = 14,15,\ j = i,\dots,16 \\
& s_{k+1}^2: & 0 &= Q(16,16) + \lambda_1
\end{aligned}
$$

The optimal values of t and λ's correspond to

$$
t = \left(\frac{L-\mu}{L+\mu}\right)^2 \qquad \lambda_1 = -\frac{2}{L+\mu} \qquad \lambda_2 = -1 \qquad \lambda_4 = \frac{L-\mu}{L+\mu}
$$

$$
\lambda_8 = 2\mu\frac{L-\mu}{(L+\mu)^2} \qquad \lambda_9 = \frac{2\mu}{L+\mu} \qquad \lambda_{10} = \left(\frac{L-\mu}{L+\mu}\right)^2 \qquad \lambda_{15} = \frac{4L\mu}{(L+\mu)^2}
$$

and $\lambda_i = 0$ for $i = 3,5,\dots,7,11,\dots,14,16$. The optimal Q is a sparse symmetric matrix, with the non-zero entries in the upper triangular portion being:

$$
Q(8,8) = \frac{2L^2\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(8,9) = -\frac{L\mu^2}{(L+\mu)^2} \qquad Q(8,10) = -\frac{L\mu^2}{(L-\mu)(L+\mu)}
$$

$$
Q(8,11) = -\frac{2L\mu^2}{(L+\mu)^2(L-\mu)} \qquad Q(8,12) = \frac{L\mu}{(L+\mu)^2} \qquad Q(8,13) = \frac{L\mu}{(L-\mu)(L+\mu)}
$$

$$
Q(8,16) = \frac{2L\mu}{(L+\mu)^2} \qquad Q(9,9) = \frac{L\mu(L+3\mu)}{2(L+\mu)^2} \qquad Q(9,10) = -\frac{L\mu}{2(L+\mu)}
$$

$$
Q(9,11) = \frac{\mu^2}{(L+\mu)^2} \qquad Q(9,12) = -\frac{\mu(3L+\mu)}{2(L+\mu)^2} \qquad Q(9,13) = -\frac{\mu}{2(L+\mu)}
$$

$$
Q(9,16) = -\frac{2L\mu}{(L+\mu)^2} \qquad Q(10,10) = \frac{L\mu}{2(L-\mu)} \qquad Q(10,11) = \frac{\mu^2}{(L-\mu)(L+\mu)}
$$

$$
Q(10,12) = \frac{\mu}{2(L+\mu)} \qquad Q(10,13) = -\frac{\mu}{2(L-\mu)} \qquad Q(11,11) = \frac{2L\mu}{(L+\mu)^2(L-\mu)}
$$

$$
Q(11,12) = -\frac{\mu}{(L+\mu)^2} \qquad Q(11,13) = -\frac{\mu}{(L-\mu)(L+\mu)} \qquad Q(12,12) = \frac{L+3\mu}{2(L+\mu)^2}
$$

$$
Q(12,13) = \frac{1}{2(L+\mu)} \qquad Q(12,16) = \frac{1}{L+\mu} \qquad Q(13,13) = \frac{1}{2(L-\mu)}
$$

$$
Q(13,16) = \frac{1}{L+\mu} \qquad Q(16,16) = \frac{2}{L+\mu}
$$

These fulfill the set of linear equalities in the SDP. Q is also PSD, as can be verified by hand or by symbolic computation. This completes the proof that

$$
f_{k+1} - f_* \le \left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*)
$$