Appendix to
Faster Randomized Infeasible Interior Point Methods for Tall/Wide Linear Programs

Appendix A Extensions

We briefly discuss extensions of our work. First, there is nothing special about using a CG solver for solving eqn. (5). We analyze two more solvers that could replace the proposed CG solver without any loss in accuracy or any increase in the number of iterations for the long-step infeasible IPM Algorithm 2 of Section 3. In Appendix D, we analyze the performance of the preconditioned Richardson Iteration and in Appendix E, we analyze the performance of the preconditioned Steepest Descent. In both cases, if the respective preconditioned solver (with the preconditioner of Section 2) runs for $t = \mathcal{O}(\log n)$ steps, Theorem 1 still holds, with small differences in the constant terms. While preconditioned Richardson iteration and preconditioned Steepest Descent are interesting from a theoretical perspective, they are not particularly practical.

Second, recall that our approach focused on full rank input matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \ll n$. Our overall approach still works if $\mathbf{A}$ in any $m \times n$ matrix that is low-rank, e.g., $\mathrm{rank}(\mathbf{A}) = k \ll \min\{m, n\}$. In that case, using the thin SVD of $\mathbf{A}$, we can rewrite the linear constraints as follows $\mathbf{U_A \Sigma_A V_A^\mathsf{T} x} = \mathbf{b}$, where $\mathbf{U_A} \in \mathbb{R}^{m \times k}$ and $\mathbf{V_A} \in \mathbb{R}^{n \times k}$ are the matrices of left and right singular vecors of $\mathbf{A}$ respectively; $\mathbf{\Sigma_A} \in \mathbb{R}^{k \times k}$ is the diagonal matrix with the k non-zero singular values of $\mathbf{A}$ as its diagonal elements. The LP of eqn. (1) can be restated as

$$\min \mathbf{c}^\mathsf{T}\mathbf{x}\,, \text{ subject to } \mathbf{V}_\mathbf{A}^\mathsf{T}\mathbf{x} = \widetilde{\mathbf{b}}\,, \mathbf{x} \geq \mathbf{0}\,, \tag{21}$$

where $\widetilde{\mathbf{b}} = \mathbf{\Sigma}_\mathbf{A}^{-1}\mathbf{U}_\mathbf{A}^\mathsf{T}\mathbf{b}$. Note that, $\mathrm{rank}(\mathbf{V_A}) = k \ll n$ and therefore eqn. (21) can be solved using our framework. The matrices $\mathbf{U_A}$, $\mathbf{V_A}$, and $\mathbf{\Sigma_A}$ can be approximately recovered using the fast SVD algorithms of [23, 5, 10]. However, the accuracy of the final solution will depend on the accuracy of the approximate SVD and we defer this analysis to future work.

Third, even though we chose to use the Count-Min sketch and its analysis from [13] (Section 1.3), there are many other alternative sketching matrix constructions that would lead to similar results. A particularly simple one is the Gaussian sketching matrix $\mathbf{W}_G \in \mathbb{R}^{n \times w}$, where every entry is a $\mathcal{N}(0,1)$ random variable. Setting $w = \mathcal{O}\left(\frac{m+\log(1/\delta)}{\zeta^2}\right)$ would result in the same accuracy guarantees as the sketching matrix of Section 1.3. However, the (theoretical) running time needed to compute $\mathbf{ADW}$ increases to $\mathcal{O}(m \cdot \mathsf{nnz}(\mathbf{A}))$. In practice, at least for relatively small matrices, using Gaussian sketching matrices is a reasonable alternative; see the discussion in [35] which argued that the Gaussian matrix sketching-based solvers are considerably better than direct solvers. We also opted to use Gaussian matrices in our empirical evaluation, since we primarily interested in measuring the accuracy of the final solution as a function of the number of iterations of the solver and the IPM algorithm. Other known constructions of sketching matrices that are also applicable in our setting include (any) sub-gaussian sketching matrix; the Subsampled Randomized Hadamard transform (SRHT); and any of the Sparse Subspace Embeddings of [9, 39, 34, 11].

We conclude by noting that our work can also be extended to analyze feasible IPMs, namely Algorithm 2 can start with a strictly feasible point. In this case, the analysis is somewhat simpler and the iteration complexity of the IPM algorithm reduces to $\mathcal{O}(n \log(1/\epsilon))$, which is the best known for feasible long-step path following IPM algorithms. We chose to present the more technically challenging infeasible IPM in this paper and delegate the feasible case to future work.

Appendix B Additional Notations

As before, we take $\mathbf{AD} = \mathbf{U\Sigma V}^\mathsf{T}$ to be the thin SVD representation of $\mathbf{AD}$. Additionally, for any two symmetric positive semidefinite (positive definite) matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ with same order, $\mathbf{A}_1 \preccurlyeq \mathbf{A}_2$ ($\mathbf{A}_1 \prec \mathbf{A}_2$) denotes that $\mathbf{A}_2 - \mathbf{A}_1$ is positive semidefinite (positive definite). For any two vectors $\mathbf{a} = (a_1, \ldots, a_\ell)^\mathsf{T}$ and $\mathbf{b} = (b_1, \ldots, b_\ell)^\mathsf{T}$ let $\mathbf{a} \circ \mathbf{b} = (a_1 b_1, \ldots, a_\ell b_\ell)^\mathsf{T}$. For any vector $\mathbf{a} \in \mathbb{R}^n$ its ℓ_∞ norm is defined as $\|\mathbf{a}\|_\infty = \max_i |a_i|$.

Appendix C Proofs

C.1 Proof of Lemma 2

Proof Consider the condition of eqn. (12):

$$\|\mathbf{V}^\top\mathbf{W}\mathbf{W}^\top\mathbf{V} - \mathbf{I}_m\|_2 \le \frac{\zeta}{2} \;\Leftrightarrow\; -\frac{\zeta}{2}\mathbf{I}_m \preccurlyeq \mathbf{V}^\top\mathbf{W}\mathbf{W}^\top\mathbf{V} - \mathbf{I}_m \preccurlyeq \frac{\zeta}{2}\mathbf{I}_m \tag{22}$$

$$\Leftrightarrow\; -\frac{\zeta}{2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top \preccurlyeq \mathbf{A}\mathbf{D}\mathbf{W}\mathbf{W}^\top\mathbf{D}\mathbf{A}^\top - \mathbf{A}\mathbf{D}^2\mathbf{A}^\top \preccurlyeq \frac{\zeta}{2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top \tag{23}$$

$$\Leftrightarrow\; \left(1-\frac{\zeta}{2}\right)\mathbf{A}\mathbf{D}^2\mathbf{A}^\top \preccurlyeq \underbrace{\mathbf{A}\mathbf{D}\mathbf{W}\mathbf{W}^\top\mathbf{D}\mathbf{A}^\top}_{\mathbf{Q}} \preccurlyeq \left(1+\frac{\zeta}{2}\right)\mathbf{A}\mathbf{D}^2\mathbf{A}^\top, \tag{24}$$

where we obtain eqn. (23) by pre- and post-multiplying the previous inequality by $\mathbf{U\Sigma}$ and $\mathbf{\Sigma U}^\top$ respectively and using the facts that $\mathbf{AD} = \mathbf{U\Sigma V}^\top$ and $\mathbf{A}\mathbf{D}^2\mathbf{A}^\top = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^\top$. Also, from eqn. (22), note that all the eigenvalues of $\mathbf{V}^\top\mathbf{W}\mathbf{W}^\top\mathbf{V}$ lie between $(1-\frac{\zeta}{2})$ and $(1+\frac{\zeta}{2})$ *i.e.*, $\mathrm{rank}(\mathbf{V}^\top\mathbf{W}) = m$. Therefore, $\mathrm{rank}(\mathbf{ADW}) = \mathrm{rank}(\mathbf{U\Sigma V}^\top\mathbf{W}) = m$, as $\mathbf{U\Sigma}$ is non-singular and we know rank of a matrix remains unaltered by pre (or post)-multiplying by a non-singular matrix. So, we have $\mathrm{rank}(\mathbf{Q}) = m$; in words $\mathbf{Q}$ has full rank. Therefore, all the diagonal entries of $\mathbf{\Sigma_Q}$ are positive and $\mathbf{Q}^{-1/2}\mathbf{Q}\mathbf{Q}^{-1/2} = (\mathbf{U_Q}\mathbf{\Sigma_Q}^{-1/2}\mathbf{U_Q}^\top)\,\mathbf{U_Q}\mathbf{\Sigma_Q}\mathbf{U_Q}^\top\,(\mathbf{U_Q}\mathbf{\Sigma_Q}^{-1/2}\mathbf{U_Q}^\top) = \mathbf{I}_m$.

Using above arguments, pre- and post- multiplying eqn. (24) by $\mathbf{Q}^{-1/2}$, we obtain

$$\left(1-\frac{\zeta}{2}\right)\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2} \preccurlyeq \mathbf{I}_m \preccurlyeq \left(1+\frac{\zeta}{2}\right)\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2}$$

$$\Leftrightarrow\; \left(1+\frac{\zeta}{2}\right)^{-1}\mathbf{I}_m \preccurlyeq \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2} \preccurlyeq \left(1-\frac{\zeta}{2}\right)^{-1}\mathbf{I}_m. \tag{25}$$

Eqn. (25) implies and is implied by the fact that all the eigenvalues of $\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2}$ are bounded between $\left(1+\frac{\zeta}{2}\right)^{-1}$ and $\left(1-\frac{\zeta}{2}\right)^{-1}$. Therefore, we have

$$\left(1+\frac{\zeta}{2}\right)^{-1} \le \sigma_i^2(\mathbf{Q}^{-1/2}\mathbf{AD}) \le \left(1-\frac{\zeta}{2}\right)^{-1}, \;\textbf{for } i = 1,\ldots,m.$$

■

C.2 Satisfying eqn. (7) using CG Solver

Let $\tilde{\mathbf{f}}^{(j)}$ be the residual at the j-th iteration of the CG algorithm, i.e., $\tilde{\mathbf{f}}^{(j)} = \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^j - \mathbf{Q}^{-1/2}\mathbf{p}$. Recall from Algorithm 1 that $\tilde{\mathbf{z}}^0 = \mathbf{0}$ and thus $\tilde{\mathbf{f}}^{(0)} = -\mathbf{Q}^{-1/2}\mathbf{p}$. In our parlance, Theorem 8 of [6] proved the following bound.

Lemma 5 (Theorem 8 of [6]) *Let $\tilde{\mathbf{f}}^{(j-1)}$ and $\tilde{\mathbf{f}}^{(j)}$ be the residuals obtained by the CG solver at steps $j-1$ and j. Then,*

$$\|\tilde{\mathbf{f}}^{(j)}\|_2 \le \frac{\kappa^2(\mathbf{Q}^{-1/2}\mathbf{AD}) - 1}{2}\|\tilde{\mathbf{f}}^{(j-1)}\|_2,$$

where $\kappa(\mathbf{Q}^{-1/2}\mathbf{AD})$ is the condition number of $\mathbf{Q}^{-1/2}\mathbf{AD}$.

From Lemma 2, we get

$$\kappa^2(\mathbf{Q}^{-1/2}\mathbf{AD}) = \frac{\sigma^2_{\max}(\mathbf{Q}^{-1/2}\mathbf{AD})}{\sigma^2_{\min}(\mathbf{Q}^{-1/2}\mathbf{AD})} \le \frac{1+\zeta/2}{1-\zeta/2}. \tag{26}$$

Combining eqn. (26) with Lemma 5,

$$\|\tilde{\mathbf{f}}^{(j)}\|_2 \le \frac{\frac{1+\zeta/2}{1-\zeta/2} - 1}{2}\|\tilde{\mathbf{f}}^{(j-1)}\|_2 = \frac{\zeta}{2-\zeta}\|\tilde{\mathbf{f}}^{(j-1)}\|_2 \le \zeta\|\tilde{\mathbf{f}}^{(j-1)}\|_2, \tag{27}$$

where the last inequality follows from $\zeta \leq 1$. Applying eqn. (27) recursively, we get

$$\|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \zeta \|\tilde{\mathbf{f}}^{(t-1)}\|_2 \leq \cdots \leq \zeta^t \|\tilde{\mathbf{f}}^{(0)}\|_2 = \zeta^t \|\mathbf{Q}^{-1/2}\mathbf{p}\|_2 \,,$$

which proves the condition of eqn. (7).

C.3 Proof of Lemma 3

Proof Let $\mathbf{AD} = \mathbf{U\Sigma V}^\mathsf{T}$ be the thin SVD representation of $\mathbf{AD}$. We use the exact same $\mathbf{W}$ as discussed in Section 2. Therefore, eqn. (12) holds with probability $1-\delta$ and it directly follows from the proof of Lemma 2 that $\operatorname{rank}(\mathbf{ADW}) = m$.

Now, as $\mathbf{ADW}$ has full *row-rank*, right-inverse exists and $\mathbf{ADW}\,(\mathbf{ADW})^\dagger = \mathbf{I}_m$. Therefore, taking $\mathbf{v} = (\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^\dagger(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p})$, we finally have

$$\begin{aligned}
\mathbf{AS}^{-1}\,\mathbf{v} &= \mathbf{AS}^{-1}(\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^\dagger(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p}) \\
&= \mathbf{ADW}(\mathbf{ADW})^\dagger(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p}) \\
&= \mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p}\,,
\end{aligned}$$

where the second equality follows from the fact that $\mathbf{D} = \mathbf{X}^{1/2}\mathbf{S}^{-1/2}$. This concludes the proof. ■

C.4 Proof of Lemma 4

Proof We already have, $\mathbf{Q} = \mathbf{ADW}(\mathbf{ADW})^\mathsf{T} = \mathbf{U_Q\Sigma_Q U_Q^\mathsf{T}}$. From this, we know that $\mathbf{U_Q}$ and $\mathbf{\Sigma_Q^{1/2}}$ are respectively the matrices of left singular vectors and singular values of $\mathbf{ADW}$. Now, let $\widehat{\mathbf{V}}$ be the right singular vector of $\mathbf{ADW}$. Therefore, $\mathbf{ADW} = \mathbf{U_Q\Sigma_Q^{1/2}}\widehat{\mathbf{V}}^\mathsf{T}$ is the thin SVD representation of $\mathbf{ADW}$. Also, from Lemma 2, we know $\mathbf{Q}$ has full rank. Therefore, $\mathbf{Q}^{1/2}\mathbf{Q}^{-1/2} = \mathbf{I}_m$.

Next, we bound $\|\mathbf{v}\|_2$ in the following way

$$\begin{aligned}
\|\mathbf{v}\|_2 &= \|(\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^\dagger(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p})\|_2 \\
&= \|(\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^\dagger\mathbf{Q}^{1/2}\mathbf{Q}^{-1/2}(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p})\|_2 \\
&\leq \|(\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^\dagger\mathbf{Q}^{1/2}\|_2\,\|\tilde{\mathbf{f}}^{(t)}\|_2\,,
\end{aligned} \tag{28}$$

where we have used the fact that $\mathbf{Q}^{-1/2}(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} - \mathbf{p}) = \tilde{\mathbf{f}}^{(t)}$ and the last inequality follows from the sub-multiplicativity property of spectral-norm.

Again, using SVD of $\mathbf{ADW}$ and $\mathbf{Q}$, we have $(\mathbf{ADW})^\dagger\mathbf{Q}^{1/2} = \widehat{\mathbf{V}}\mathbf{\Sigma_Q^{-1/2}U_Q^\mathsf{T}}\,\mathbf{U_Q\Sigma_Q^{1/2}U_Q^\mathsf{T}} = \widehat{\mathbf{V}}\mathbf{U_Q^\mathsf{T}}$. Now, note that $\mathbf{U_Q} \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and $\widehat{\mathbf{V}} \in \mathbb{R}^{w\times m}$ has orthogonal columns *i.e.* $\|\widehat{\mathbf{V}}\|_2 = 1$. Therefore, combining these with eqn. (28) yields,

$$\begin{aligned}
\|\mathbf{v}\|_2 &\leq \|(\mathbf{XS})^{1/2}\mathbf{W}\widehat{\mathbf{V}}\mathbf{U_Q^\mathsf{T}}\|_2\|\tilde{\mathbf{f}}^{(t)}\|_2 = \|(\mathbf{XS})^{1/2}\mathbf{W}\widehat{\mathbf{V}}\|_2\|\tilde{\mathbf{f}}^{(t)}\|_2 \\
&\leq \|(\mathbf{XS})^{1/2}\mathbf{W}\|_2\|\widehat{\mathbf{V}}\|_2\|\tilde{\mathbf{f}}^{(t)}\|_2 = \|(\mathbf{XS})^{1/2}\mathbf{W}\|_2\|\tilde{\mathbf{f}}^{(t)}\|_2\,,
\end{aligned} \tag{29}$$

where the first equality in eqn. (29) follows from the unitary invariance property of the spectral norm; the second inequality follows from the sub-multiplicativity of the spectral norm and the last equality is due to $\|\widehat{\mathbf{V}}\|_2 = 1$. Now, as we use the exact same $\mathbf{W}$ discussed in Section 2 to construct $\mathbf{v}$ and note that eqn. (10) holds for any matrix $\mathbf{Z}$ (irrespective of its dimensions). Therefore, taking $\mathbf{Z} = (\mathbf{XS})^{1/2}$ with that $\mathbf{W}$, eqn. (10) in Section 1.3 boils down to

$$\left\|(\mathbf{XS})^{1/2}\mathbf{WW}^\mathsf{T}(\mathbf{XS})^{1/2} - (\mathbf{XS})\right\|_2 \leq \frac{\zeta}{4}\left(\|(\mathbf{XS})^{1/2}\|_2^2 + \frac{\|(\mathbf{XS})^{1/2}\|_F^2}{m}\right) \tag{30}$$

holds with probability at least $1-\delta$.

Now, applying Weyl's inequality on the left hand side of the eqn. (30), we further have

$$\left|\left\|(\mathbf{XS})^{1/2}\mathbf{W}\right\|_2^2 - \left\|(\mathbf{XS})^{1/2}\right\|_2^2\right| \leq \frac{\zeta}{4}\left(\|(\mathbf{XS})^{1/2}\|_2^2 + \frac{\|(\mathbf{XS})^{1/2}\|_F^2}{m}\right) \tag{31}$$

Now, using the facts that $\frac{\zeta}{4} \leq 1$, $\|(\mathbf{XS})^{1/2}\|_2 \leq \|(\mathbf{XS})^{1/2}\|_F$, and $\frac{\|(\mathbf{XS})^{1/2}\|_F^2}{m} \leq \|(\mathbf{XS})^{1/2}\|_F^2$, from eqn. (31),

$$\left\|(\mathbf{XS})^{1/2}\mathbf{W}\right\|_2^2 \leq 3\|(\mathbf{XS})^{1/2}\|_F^2 = 3n\mu\,, \tag{32}$$

where the last equality follows from $\|(\mathbf{XS})^{1/2}\|_F^2 = \mathbf{x}^\mathsf{T}\mathbf{s} = n\mu$.

Finally, combining eqns. (29) and (32), we conclude

$$\|\mathbf{v}\|_2 \leq \sqrt{3n\mu}\|\tilde{\mathbf{f}}^{(t)}\|_2.$$

■

Appendix D Richardson Iteration

Here, we show that all our analyses still hold, even if we replace Step 3 of Algorithm 1 (CG solver) with Richardson iteration. Basically, all we need to show is the condition in eqn. (7) holds. Note that the condition in eqn. (6) already holds from Lemma 2, as we use the exact same sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$ discussed in Section 2.

Algorithm 3 Richardson Iteration Solver

Input: $\mathbf{AD} \in \mathbb{R}^{m\times n}$, $\mathbf{p} \in \mathbb{R}^m$; number of iterations $t > 0$; sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$;
Initialize: $\tilde{\mathbf{z}}^0 \leftarrow \mathbf{0}_m$;
for $j = 1$ **to** t **do**
  $\tilde{\mathbf{z}}^j \leftarrow \tilde{\mathbf{z}}^{j-1} + \mathbf{Q}^{-1/2}(\mathbf{p} - \mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j-1})$;
end for
Output: return $\tilde{\mathbf{z}}^t$;

Our first result expresses the residual vector $\tilde{\mathbf{f}}^{(j)}$ in terms of $\tilde{\mathbf{f}}^{(j-1)}$ for $j = 1, 2, \ldots, t$.

Lemma 6 *Let $\tilde{\mathbf{f}}^{(j)}$, $j = 1, 2, \ldots, t$ be the residual vectors at each iteration.Then,*

$$\tilde{\mathbf{f}}^{(j)} = \left(\mathbf{I}_m - \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right)\tilde{\mathbf{f}}^{(j-1)}\,. \tag{33}$$

Recall that $\mathbf{Q} = \mathbf{ADWW}^\mathsf{T}\mathbf{DA}^\mathsf{T}$ and $\tilde{\mathbf{f}}^{(j)} = \mathbf{Q}^{-1/2}(\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^j - \mathbf{p})$.

Proof Using Algorithm 3, we express $\tilde{\mathbf{f}}^{(j)}$ as

$$\begin{aligned}
\tilde{\mathbf{f}}^{(j)} &= \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^j - \mathbf{Q}^{-1/2}\mathbf{p} \\
&= \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\left(\tilde{\mathbf{z}}^{j-1} + \mathbf{Q}^{-1/2}(\mathbf{p} - \mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j-1})\right) - \mathbf{Q}^{-1/2}\mathbf{p} \\
&= \left(\mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j-1} - \mathbf{Q}^{-1/2}\mathbf{p}\right) \\
&\qquad\qquad - \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\left(\mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j-1} - \mathbf{Q}^{-1/2}\mathbf{p}\right) \\
&= \left(\mathbf{I}_m - \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right)\left(\mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j-1} - \mathbf{Q}^{-1/2}\mathbf{p}\right) \\
&= \left(\mathbf{I}_m - \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right)\tilde{\mathbf{f}}^{(j-1)}\,,
\end{aligned}$$

which concludes the proof. ■

In the next result, we show that the spectral norm of $(\mathbf{I}_m - \mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2})$ is upper bounded by ζ.

Lemma 7 *Let the condition in eqn. (6) holds for the sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$, then*

$$\|\mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2} - \mathbf{I}_m\|_2 \leq \zeta\,.$$

Proof As the condition in eqn. (6) holds, we can go backwards in the proof of Lemma 2 and see that eqn. (25) holds. So, we subtract $\mathbf{I}_m$ from each side of eqn. (25) to get

$$\begin{aligned}
&\left(\frac{2}{2+\zeta}-1\right)\mathbf{I}_m \preccurlyeq \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \left(\frac{2}{2-\zeta}-1\right)\mathbf{I}_m \\
\Leftrightarrow\ & -\frac{\zeta}{2+\zeta}\mathbf{I}_m \preccurlyeq \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \frac{\zeta}{2-\zeta}\mathbf{I}_m \\
\Rightarrow\ & -\frac{\zeta}{2-\zeta}\mathbf{I}_m \preccurlyeq \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \frac{\zeta}{2-\zeta}\mathbf{I}_m \qquad (34)\\
\Leftrightarrow\ & \|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m\|_2 \le \frac{\zeta}{2-\zeta} \le \zeta\,, \qquad (35)
\end{aligned}$$

where eqn. (34) holds as $\frac{\zeta}{2+\zeta} \le \frac{\zeta}{2-\zeta}$ and the last inequality of eqn. (35) follows from $\zeta < 1$. ■

Satisfying eqn. (6). Note that the condition in eqn. (6) already holds from Lemma 2, as we use the exact same sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$ discussed in Section 2.

Satisfying eqn. (7). Using Lemma 7 and applying Lemma 6 recursively, we get

$$\|\tilde{\mathbf{f}}^{(t)}\|_2 \le \zeta\|\tilde{\mathbf{f}}^{(t-1)}\|_2 \le \cdots \le \zeta^t\|\tilde{\mathbf{f}}^{(0)}\|_2 = \zeta^t\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2\,.$$

Appendix E Steepest Descent

We will now replace Step 3 of Algorithm 1 (our proposed CG solver) by preconditioned steepest descent. We will again prove that our analysis of the proposed infeasible long-step IPM remains essentially the same.

First, we construct the sketching matrix $\mathbf{W}$ as discussed in Section 1.3, with a slightly more stringent accuracy guarantee. More specifically, we necessitate that

$$\left\|\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{W}^\mathsf{T}\mathbf{V}-\mathbf{I}_m\right\|_2 \le \frac{\zeta(1-\zeta)}{2} \qquad (36)$$

holds with probability at least $1-\delta$ for a constant $\zeta \in [0,1]$. Notice that the sketching dimension $w = \mathcal{O}(m\log(m/\delta))$ and the running time needed to compute $\mathbf{Q}^{-1/2}$ (which is $\mathcal{O}(\mathsf{nnz}(\mathbf{A})\cdot\log(m/\delta)+m^3\log(m/\delta))$) remain, asymptotically, the same. In the case of steepest descent, it turns out that at each iteration the search direction is the negative of the gradient, which is equal to the residual $\tilde{\mathbf{f}}^{(j)}$. Moreover, the step size α_j is determined by an exact *line search* that minimizes the underlying quadratic function:

$$\alpha_j = \frac{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\tilde{\mathbf{f}}^{(j)}}{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)}}.$$

For this choice of α_j, it is easy to verify that the current gradient is orthogonal to the previous one.

Algorithm 4 Steepest Descent Solver

Input: $\mathbf{A}\mathbf{D} \in \mathbb{R}^{m\times n}$, $\mathbf{p} \in \mathbb{R}^m$; number of iterations $t > 0$; sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$;
Initialize: $\tilde{\mathbf{z}}^0 \leftarrow \mathbf{0}_m$;
for $j = 0$ **to** $t-1$ **do**

  $\alpha_j = \frac{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\tilde{\mathbf{f}}^{(j)}}{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)}}$;

  $\tilde{\mathbf{z}}^{j+1} \leftarrow \tilde{\mathbf{z}}^j - \alpha_j\tilde{\mathbf{f}}^{(j)}$;
end for

Output: return = $\tilde{\mathbf{z}}^t$;

Similar to Lemma 6, our next result reveals a recursive relation between the search directions which, later on, will be instrumental in bounding $\tilde{\mathbf{f}}^{(t)}$.

Lemma 8 *Let $\tilde{\mathbf{f}}^{(j)}$, $j = 1, 2, \ldots, t$ be the residual vectors at each iteration and α_j is given by Algorithm 4. Then,*

$$\tilde{\mathbf{f}}^{(j+1)} = \left(\mathbf{I}_m - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right)\tilde{\mathbf{f}}^{(j)}, \tag{37}$$

Recall that $\mathbf{Q} = \mathbf{A}\mathbf{D}\mathbf{W}\mathbf{W}^\mathsf{T}\mathbf{D}\mathbf{A}^\mathsf{T}$ and $\tilde{\mathbf{f}}^{(j)} = \mathbf{Q}^{-1/2}(\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^j - \mathbf{p})$.

Proof In Algorithm 4, we pre-multiply $\tilde{\mathbf{z}}^{j+1}$ with $\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}$ and then subtract $\mathbf{Q}^{-1/2}\mathbf{p}$ to get

$$\begin{aligned}
\tilde{\mathbf{f}}^{(j+1)} &= \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j+1} - \mathbf{Q}^{-1/2}\mathbf{p} \\
&= \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^{j} - \mathbf{Q}^{-1/2}\mathbf{p} - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)} \\
&= \tilde{\mathbf{f}}^{(j)} - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)} = \left(\mathbf{I}_m - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right)\tilde{\mathbf{f}}^{(j)},
\end{aligned}$$

which concludes the proof. ■

Next, using this new condition in eqn. (36), we will bound $\left\|\mathbf{I}_m - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right\|_2$ through a couple of results.

Lemma 9 *If eqn. (36) is satisfied, then $|\alpha_j - 1| \le \frac{\zeta(1-\zeta)}{2}$.*

Proof First, we rewrite eqn. (36) as follows,

$$-\frac{\zeta(1-\zeta)}{2}\mathbf{I}_m \preccurlyeq \mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{W}^\mathsf{T}\mathbf{V} - \mathbf{I}_m \preccurlyeq \frac{\zeta(1-\zeta)}{2}\mathbf{I}_m$$

Next, we pre and post-multiply the the above expression by $\mathbf{U}\mathbf{\Sigma}$ and $\mathbf{\Sigma}\mathbf{U}^\mathsf{T}$ to get

$$-\frac{\zeta(1-\zeta)}{2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T} \preccurlyeq \underbrace{\mathbf{A}\mathbf{D}\mathbf{W}\mathbf{W}^\mathsf{T}\mathbf{D}\mathbf{A}^\mathsf{T}}_{\mathbf{Q}} - \mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T} \preccurlyeq \frac{\zeta(1-\zeta)}{2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T} \tag{38}$$

Now, pre and post-multiplying eqn. (38) again by $\mathbf{Q}^{-1/2}$, we have

$$\begin{aligned}
&\left(1 - \frac{\zeta(1-\zeta)}{2}\right)\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2} \preccurlyeq \mathbf{I}_m \preccurlyeq \left(1 + \frac{\zeta(1-\zeta)}{2}\right)\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2} \\
&\Rightarrow \left(1 - \frac{\zeta(1-\zeta)}{2}\right)\tilde{\mathbf{f}}^{(j)\mathsf{T}}\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)} \le \tilde{\mathbf{f}}^{(j)\mathsf{T}}\tilde{\mathbf{f}}^{(j)} \le \left(1 + \frac{\zeta(1-\zeta)}{2}\right)\tilde{\mathbf{f}}^{(j)\mathsf{T}}\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)} \\
&\Rightarrow \left(1 - \frac{\zeta(1-\zeta)}{2}\right) \le \frac{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\tilde{\mathbf{f}}^{(j)}}{\tilde{\mathbf{f}}^{(j)\mathsf{T}}\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\tilde{\mathbf{f}}^{(j)}} \le \left(1 + \frac{\zeta(1-\zeta)}{2}\right) \\
&\Leftrightarrow |\alpha_j - 1| \le \frac{\zeta(1-\zeta)}{2}, \text{ for } j = 1, 2, \ldots, t.
\end{aligned} \tag{39}$$

■

Our next result shows that under eqn. (36), $\left\|\mathbf{I}_m - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right\|_2$ is upper bounded by a small quantity for for $j = 1, 2, \ldots, t$.

Lemma 10 *If eqn. (36) is satisfied, then $\left\|\mathbf{I}_m - \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right\|_2 \le \zeta$, for $j = 1, 2, \ldots, t$.*

Proof We note that eqn. (36) directly implies

$$\left\|\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{W}^\mathsf{T}\mathbf{V} - \mathbf{I}_m\right\|_2 \le \frac{\zeta}{2} \tag{40}$$

Now, as eqn. (40) holds, from eqn. (25) in the proof of Lemma 2, we have

$$\left(1 + \frac{\zeta}{2}\right)^{-1}\mathbf{I}_m \preccurlyeq \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2} \preccurlyeq \left(1 - \frac{\zeta}{2}\right)^{-1}\mathbf{I}_m$$

$$\Leftrightarrow \left(\frac{2\alpha_j}{2+\zeta}-1\right)\mathbf{I}_m \preccurlyeq \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \left(\frac{2\alpha_j}{2-\zeta}-1\right)\mathbf{I}_m$$

$$\Leftrightarrow \frac{2(\alpha_j-1)-\zeta}{2+\zeta}\,\mathbf{I}_m \preccurlyeq \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \frac{2(\alpha_j-1)+\zeta}{2-\zeta}\,\mathbf{I}_m\,, \tag{41}$$

where the above expression follows from multiplying eqn. (25) by α_j and then subtracting $\mathbf{I}_m$.

Now, from Lemma 9, we have, $-\zeta(1-\zeta) \le 2(\alpha_j-1) \le \zeta(1-\zeta)$ for $j=1,2,\ldots,t$. Using this in eqn. (41), we further have

$$-\frac{\zeta(1-\zeta)+\zeta}{2+\zeta}\,\mathbf{I}_m \preccurlyeq \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \frac{\zeta(1-\zeta)+\zeta}{2-\zeta}\,\mathbf{I}_m$$

$$\Leftrightarrow -\frac{\zeta(2-\zeta)}{2+\zeta}\,\mathbf{I}_m \preccurlyeq \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \zeta\,\mathbf{I}_m$$

$$\Rightarrow -\zeta\,\mathbf{I}_m \preccurlyeq \alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}-\mathbf{I}_m \preccurlyeq \zeta\,\mathbf{I}_m \tag{42}$$

$$\Rightarrow \left\|\mathbf{I}_m-\alpha_j \mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2}\right\|_2 \le \zeta\,,$$

where eqn. (42) is due to the fact that $\frac{2-\zeta}{2+\zeta} \le 1$. ■

Satisfying eqn. (6). As eqn. (40) holds, eqn. (6) directly follows from Lemma 2.

Satisfying eqn. (7). Using Lemma 10 and applying Lemma 8 recursively, we get

$$\|\tilde{\mathbf{f}}^{(t)}\|_2 \le \zeta\|\tilde{\mathbf{f}}^{(t-1)}\|_2 \le \cdots \le \zeta^t\|\tilde{\mathbf{f}}^{(0)}\|_2 = \zeta^t\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2\,.$$

Appendix F Convergence Analysis of Algorithm 2

F.1 Number of Iterations for the CG Solver

In this section, most of the proofs follow [38] except for the fact that we used our sketching based preconditioner $\mathbf{Q}^{-1/2}$. Recall that $\mathcal{S}$ is the set of optimal and feasible solutions for the proposed LP.

Lemma 11 *Let $(\mathbf{x}^0,\mathbf{y}^0,\mathbf{s}^0)$ be the initial point with $(\mathbf{x}^0,\mathbf{s}^0) > \mathbf{0}$ and $(\mathbf{x}^*,\mathbf{y}^*,\mathbf{s}^*) \in \mathcal{S}$ such that $(\mathbf{x}^*,\mathbf{s}^*) \le (\mathbf{x}^0,\mathbf{s}^0)$ with $\mathbf{s}^0 \ge |\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c}|$. Then, for any point $(\mathbf{x},\mathbf{y},\mathbf{s}) \in \mathcal{N}(\gamma)$ such that $\mathbf{r}=\eta\,\mathbf{r}^0$ and $0 \le \eta \le \min\left\{1, \frac{\mathbf{s}^\mathsf{T}\mathbf{x}}{\mathbf{s}^{0\mathsf{T}}\mathbf{x}^0}\right\}$, then we have*

$$(i)\;\; \eta\,(\mathbf{x}^\mathsf{T}\mathbf{s}^0+\mathbf{s}^\mathsf{T}\mathbf{x}^0) \le\; 3n\mu\,, \tag{43a}$$

$$(ii)\;\; \eta\,\|\mathbf{S}(\mathbf{x}^*-\mathbf{x}^0)\|_2 \le \eta\,\|\mathbf{S}\mathbf{x}^0\|_2 \le \eta\mathbf{s}^\mathsf{T}\mathbf{x}^0 \le\; 3n\mu\,, \tag{43b}$$

$$(iii)\;\; \eta\,\|\mathbf{X}(\mathbf{s}^0+\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})\|_2 \le\; 2\eta\,\|\mathbf{X}\mathbf{s}^0\|_2 \le\; 2\eta\,\mathbf{x}^\mathsf{T}\mathbf{s}^0 \le\; 6n\mu\,. \tag{43c}$$

Proof We prove eqns. (43a)–(43c) below.

Proof of eqn. (43a). For completeness, we provide a proof of eqn. (43a) which is already discussed in [38]. Since $(\mathbf{x}^*,\mathbf{s}^*,\mathbf{y}^*) \in \mathcal{S}$, the following equalities hold:

$$\mathbf{A}\mathbf{x}^* = \mathbf{b} \tag{44a}$$

$$\mathbf{A}^\mathsf{T}\mathbf{y}^*+\mathbf{s}^* = \mathbf{c} \tag{44b}$$

Furthermore, $\mathbf{r}=\eta\mathbf{r}^0$ implies

$$\mathbf{A}\mathbf{x}-\mathbf{b} = \eta(\mathbf{A}\mathbf{x}^0-b) \tag{45a}$$

$$\mathbf{A}^\mathsf{T}\mathbf{y}+\mathbf{s}-\mathbf{c} = \eta(\mathbf{A}^\mathsf{T}\mathbf{y}^0+\mathbf{s}^0-\mathbf{c}) \tag{45b}$$

Combining eqn. (44a) with eqn. (45a) and eqn. (44b) with eqn. (45b), we get

$$\mathbf{A}\big(\mathbf{x}-\eta\mathbf{x}^0-(1-\eta)\mathbf{x}^*\big) = \mathbf{0} \tag{46a}$$

$$\mathbf{A}^\mathsf{T}(\mathbf{y}-\eta\mathbf{y}^0-(1-\eta)\mathbf{y}^*)+(\mathbf{s}-\eta\mathbf{s}^0-(1-\eta)\mathbf{s}^*)=\mathbf{0} \tag{46b}$$

Multiplying the eqn. (46b) by $\left(\mathbf{x}-\eta\mathbf{x}^0-(1-\eta)\mathbf{x}^*\right)^\mathsf{T}$ on the left and using eqn. (46a), we get

$$\left(\mathbf{x}-\eta\mathbf{x}^0-(1-\eta)\mathbf{x}^*\right)^\mathsf{T}\left(\mathbf{s}-\eta\mathbf{s}^0-(1-\eta)\mathbf{s}^*\right)=0\,,$$

expanding which we get

$$\begin{aligned}\eta\left(\mathbf{x}^{0^\mathsf{T}}\mathbf{s}+\mathbf{x}^\mathsf{T}\mathbf{s}^0\right)&=\eta^2\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0+(1-\eta)^2(\mathbf{x}^*)^\mathsf{T}\mathbf{s}^*+\mathbf{x}^\mathsf{T}\mathbf{s}\\&\quad+\eta(1-\eta)\left(\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^*+(\mathbf{x}^*)^\mathsf{T}\mathbf{s}^0\right)-(1-\eta)\left((\mathbf{x}^*)^\mathsf{T}\mathbf{s}+\mathbf{x}^\mathsf{T}\mathbf{s}^*\right)\end{aligned} \tag{47}$$

Next, we use the given conditions and rewrite eqn. (47) as

$$\begin{aligned}\eta\left(\mathbf{x}^{0^\mathsf{T}}\mathbf{s}+\mathbf{s}^{0^\mathsf{T}}\mathbf{x}\right)&\le\eta^2\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0+\mathbf{x}^\mathsf{T}\mathbf{s}+\eta(1-\eta)\left(\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^*+\mathbf{s}^{0^\mathsf{T}}\mathbf{x}^*\right)\\&\le\eta^2\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0+\mathbf{x}^\mathsf{T}\mathbf{s}+2\eta(1-\eta)\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0\\&\le2\eta\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0+\mathbf{x}^\mathsf{T}\mathbf{s}\le3\mathbf{x}^\mathsf{T}\mathbf{s}=3n\mu\,,\end{aligned} \tag{48}$$

where the first inequality in eqn. (48) follows from from a couple of facts. First, $(1-\eta)((\mathbf{x}^*)^\mathsf{T}\mathbf{s}+\mathbf{x}^\mathsf{T}\mathbf{s}^*)\ge0$ as $(\mathbf{x}^*,\mathbf{s}^*)\ge\mathbf{0}$ and $(\mathbf{x}^0,\mathbf{s}^0)\ge\mathbf{0}$; second, as $(\mathbf{x}^*,\mathbf{s}^*,\mathbf{y}^*)\in\mathcal{S}$ (which implies $\mathbf{x}^*\circ\mathbf{s}^*=\mathbf{0}$), we have $(\mathbf{x}^*)^\mathsf{T}\mathbf{s}^*=0$. Second inequality in eqn. (48) holds as $\mathbf{x}^*\le\mathbf{x}^0$, $\mathbf{s}^*\le\mathbf{s}^0$, $(\mathbf{x}^*,\mathbf{s}^*)\ge\mathbf{0}$ and $(\mathbf{x}^0,\mathbf{s}^0)\ge\mathbf{0}$; combining which we have $(\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^*+\mathbf{s}^{0^\mathsf{T}}\mathbf{x}^*)\le2\,\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0$. Third inequality in eqn. (48) is true as we have $\eta^2\mathbf{x}^{0^\mathsf{T}}+2\eta(1-\eta)\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0=2\eta\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0-\eta^2\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0\le2\eta\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0$. Final inequality holds as $\eta\le\frac{\mathbf{x}^\mathsf{T}\mathbf{s}}{\mathbf{x}^{0^\mathsf{T}}\mathbf{s}^0}$.

Proof of eqn. (43b)**.** The last inequality directly follows from eqn. (43a); second last inequality is also easy to prove as

$$\|\mathbf{S}\mathbf{x}^0\|_2=\sqrt{\sum_{i=1}^{s}(s_ix_i^0)^2}\le\sqrt{\left(\sum_{i=1}^{s}s_ix_i^0\right)^2}=\mathbf{s}^\mathsf{T}\mathbf{x}^0\,. \tag{49}$$

To prove the first inequality in eqn. (43b), we use the fact $\mathbf{x}^0\ge\mathbf{x}^*$ as follows

$$\begin{aligned}\|\mathbf{S}\mathbf{x}^0\|_2^2-\|\mathbf{S}(\mathbf{x}^*-\mathbf{x}^0)\|_2^2&=\sum_{i=1}^{n}(s_ix_i^0)^2-\sum_{i=1}^{n}s_i^2\left((x_i^*)^2+(x_i^0)^2-2x_i^*x_i^0\right)\\&=\sum_{i=1}^{n}s_i^2\left(2x_i^*x_i^0-(x_i^*)^2\right)\ge0\,.\end{aligned}$$

Proof of eqn. (43c)**.** This can be proven using a similar approach as in eqn. (43b). Last inequality directly follows from eqn. (43a); second last inequality is also easy to prove as

$$\|\mathbf{X}\mathbf{s}^0\|_2=\sqrt{\sum_{i=1}^{n}(x_is_i^0)^2}\le\sqrt{\left(\sum_{i=1}^{n}x_is_i^0\right)^2}=\mathbf{x}^\mathsf{T}\mathbf{s}^0\,. \tag{50}$$

For the first inequality, we proceed as follows

$$\begin{aligned}\|\mathbf{X}(\mathbf{s}^0+\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})\|_2^2&=\|\mathbf{X}\mathbf{s}^0\|_2^2+\|\mathbf{X}(\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})\|_2^2+2\mathbf{s}^{0^\mathsf{T}}\mathbf{X}^\mathsf{T}\mathbf{X}(\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})\\&=\|\mathbf{X}\mathbf{s}^0\|_2^2+\sum_{i=1}^{n}x_i^2(\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})_i^2+2\sum_{i=1}^{n}x_i^2s_i^0(\mathbf{A}^\mathsf{T}\mathbf{y}^0-\mathbf{c})_i\\&\le\|\mathbf{X}\mathbf{s}^0\|_2^2+\sum_{i=1}^{n}(x_is_i^0)^2+2\sum_{i=1}^{n}(x_is_i^0)^2\end{aligned}$$

$$= \|\mathbf{X}\mathbf{s}^0\|_2^2 + \|\mathbf{X}\mathbf{s}^0\|_2^2 + 2\|\mathbf{X}\mathbf{s}^0\|_2^2 = 4\|\mathbf{X}\mathbf{s}^0\|_2^2\,, \tag{51}$$

where the inequality in eqn. (51) follows from $x_i \geq 0$, $s_i^0 \geq 0$ and $\left|(\mathbf{A}^\mathsf{T}\mathbf{y}^0 - \mathbf{c})_i\right| \leq s_i^0$ for all $i = 1, 2, \ldots n$. This concludes the proof of Lemma 11. ■

Our next result bounds $\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2$ which will be instrumental in proving the final convergence bound.

Lemma 12 *Let $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0)$ be the initial point with $(\mathbf{x}^0, \mathbf{s}^0) > \mathbf{0}$ such that $\mathbf{x}^0 \geq \mathbf{x}^*$ and $\mathbf{s}^0 \geq \max\{\mathbf{s}^*, |\mathbf{c} - \mathbf{A}^\mathsf{T}\mathbf{y}^0|\}$ for some $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*) \in \mathcal{S}$. Furthermore, let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$ with $\mathbf{r} = \eta\,\mathbf{r}^0$ for some $0 \leq \eta \leq 1$. If the sketching matrix $\mathbf{W} \in \mathbb{R}^{n \times w}$ satisfies the condition in eqn. (6), then*

$$\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2 \leq \sqrt{2}\left(\frac{9n}{\sqrt{1-\gamma}} + \sigma\sqrt{\frac{n}{1-\gamma}} + \sqrt{n}\right)\sqrt{\mu}\,.$$

Recall that, $\mathbf{r} = (\mathbf{r}_p, \mathbf{r}_d) = (\mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} - \mathbf{c})$ and $\mathbf{r}^0 = (\mathbf{r}_p^0, \mathbf{r}_d^0) = (\mathbf{A}\mathbf{x}^0 - \mathbf{b}, \mathbf{A}^\mathsf{T}\mathbf{y}^0 + \mathbf{s}^0 - \mathbf{c})$.

Proof Note that after correcting the approximation error of the CG solver using $\mathbf{v}$, the primal and dual residuals $\mathbf{r} = (\mathbf{r}_p, \mathbf{r}_d)$ corresponding to an iterate $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$ always lie on the line segment between zero and $\mathbf{r}^{(0)}$. In other words, $\mathbf{r} = \eta\mathbf{r}^{(0)}$ always holds for some $\eta \in [0, 1]$. This was formally proven in Lemma 3.3 of [38]. To bound $\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2$, first we express $\mathbf{p}$ as in eqn. (3) and rewrite

$$\mathbf{Q}^{-1/2}\mathbf{p} = \mathbf{Q}^{-1/2}\left(-\mathbf{r}_p - \sigma\mu\mathbf{A}\mathbf{S}^{-1}\mathbf{1}_n + \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{D}^2\mathbf{r}_d\right) \tag{52}$$

Then, applying triangle inequality on $\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2$ in eqn. (52), we get

$$\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2 \leq \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4\,, \tag{53}$$

where

$$\begin{aligned}
\Delta_1 &= \|\mathbf{Q}^{-1/2}\mathbf{r}_p\|_2\,,\\
\Delta_2 &= \sigma\mu\|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{1}_n\|_2\,,\\
\Delta_3 &= \|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}\mathbf{D}^{-1}\mathbf{x}\|_2\,,\\
\Delta_4 &= \|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{r}_d\|_2\,.
\end{aligned}$$

To bound Δ_1, Δ_2, Δ_3 and Δ_4 separately, we will heavily use the condition in eqn. (6). In particular, from eqn. (6), note that we have $\|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}\|_2 \leq \sqrt{2}$ as $\zeta \leq 1$.

Bounding Δ_1. Putting $\mathbf{r}_p = \eta\,\mathbf{r}_p^0$, $\mathbf{r}_p^0 = \mathbf{A}\mathbf{x}^0 - \mathbf{b}$ and $\mathbf{b} = \mathbf{A}\mathbf{x}^*$, we rewrite Δ_1 as

$$\begin{aligned}
\Delta_1 &= \eta\,\|\mathbf{Q}^{-1/2}\mathbf{A}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\\
&= \eta\,\|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}\mathbf{D}^{-1}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\\
&\leq \eta\,\|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}\|_2\|\mathbf{D}^{-1}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\\
&\leq \sqrt{2}\eta\,\|\mathbf{D}^{-1}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\\
&= \sqrt{2}\eta\,\|(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{S}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\\
&\leq \sqrt{2}\eta\,\|(\mathbf{X}\mathbf{S})^{-1/2}\|_2\,\|\mathbf{S}(\mathbf{x}^0 - \mathbf{x}^*)\|_2\,,
\end{aligned} \tag{54}$$

where the above steps follow from submultiplicativity and eqn. (6). From eqn. (6), note that we have $\|\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}\|_2 \leq \sqrt{2}$ as $\zeta \leq 1$. Now, applying eqn. (43b) and $\|(\mathbf{X}\mathbf{S})^{-1/2}\|_2 = \max_{1\leq i\leq n}\frac{1}{\sqrt{x_i s_i}}$, we further have

$$\begin{aligned}
\Delta_1 &\leq \sqrt{2}\max_{1\leq i\leq n}\frac{1}{\sqrt{x_i s_i}}\cdot 3n\mu\\
&\leq 3\sqrt{2}\,n\sqrt{\frac{\mu}{1-\gamma}}\,,
\end{aligned} \tag{55}$$

where the last inequality follows from $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$.

Bounding Δ_2. Applying submultiplicativity, we have

$$
\begin{aligned}
\Delta_2 &= \sigma\mu \,\|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\,(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{1}_n\|_2 \\
&\leq \sigma\mu \,\|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\|_2\|(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{1}_n\|_2 \\
&\leq \sqrt{2}\,\sigma\mu \,\|(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{1}_n\|_2 \\
&= \sqrt{2}\,\sigma\mu \sqrt{\sum_{i=1}^{n}\frac{1}{x_i s_i}} \leq \sqrt{2}\,\sigma\mu\sqrt{\sum_{i=1}^{n}\frac{1}{(1-\gamma)\mu}} \\
&= \sqrt{2}\,\sigma\sqrt{\frac{n\,\mu}{(1-\gamma)}}\,,
\end{aligned}
\tag{56}
$$

where the second last inequality follows from eqn. (6) and last inequality holds as $(\mathbf{x},\mathbf{y},\mathbf{s})\in\mathcal{N}(\gamma)$.

Bounding Δ_3. Putting $\mathbf{D}=\mathbf{S}^{-1/2}\mathbf{X}^{1/2}$; $\mathbf{x}=\mathbf{X}\,\mathbf{1}_n$ and

$$
\begin{aligned}
\Delta_3 &= \|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\,(\mathbf{S}^{1/2}\mathbf{X}^{-1/2})\,\mathbf{X}\,\mathbf{1}_n\|_2 \\
&= \|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\,(\mathbf{S}\mathbf{X})^{1/2}\,\mathbf{1}_n\|_2 \\
&\leq \|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\|_2\|(\mathbf{S}\mathbf{X})^{1/2}\,\mathbf{1}_n\|_2 \\
&\leq \sqrt{2}\sqrt{\sum_{i=1}^{n}x_i s_i} = \sqrt{2n\,\mu}\,,
\end{aligned}
\tag{57}
$$

where the inequalities follows respectively from submultiplicativity and eqn. (6).

Bounding Δ_4. Putting $\mathbf{r}_d=\eta\,\mathbf{r}_d^0$, we have

$$
\begin{aligned}
\Delta_4 &= \eta\|\mathbf{Q}^{-1/2}\,\mathbf{A}\,\mathbf{D}^2\mathbf{r}_d^0\|_2 \\
&\leq \eta\|\mathbf{Q}^{-1/2}\,\mathbf{A}\mathbf{D}\|_2\|(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{X}\mathbf{r}_d^0\|_2 \\
&\leq \sqrt{2}\eta\,\|(\mathbf{X}\mathbf{S})^{-1/2}\mathbf{X}(\mathbf{A}^\top\mathbf{y}^0+\mathbf{s}^0-\mathbf{c})\|_2 \\
&\leq \sqrt{2}\eta\,\|(\mathbf{X}\mathbf{S})^{-1/2}\|_2\,\|\mathbf{X}(\mathbf{A}^\top\mathbf{y}^0+\mathbf{s}^0-\mathbf{c})\|_2\,,
\end{aligned}
$$

where the above inequalities follow from submultiplicativity and eqn. (6). Now, applying eqn. (43c) and $\|(\mathbf{X}\mathbf{S})^{-1/2}\|_2\leq\frac{1}{\sqrt{(1-\gamma)\mu}}$, we further have

$$
\Delta_4 \leq\; 6\sqrt{2}n\sqrt{\frac{\mu}{1-\gamma}}
\tag{58}
$$

Final bound. Combining eqns. (53), (55) ,(56), (57) and (58)

$$
\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2 \leq\; \sqrt{2}\left(\frac{9n}{\sqrt{1-\gamma}}+\sigma\sqrt{\frac{n}{1-\gamma}}+\sqrt{n}\right)\sqrt{\mu}\,.
\tag{59}
$$

This concludes the proof of Lemma 12. ■

Lemma 13 *Let the sketching matrix* $\mathbf{W}$ *satisfy the conditions in eqns. (6) and (7). Then, after* $t\geq\frac{\log(4\sqrt{6n}\,\psi/\gamma\sigma)}{\log(1/\zeta)}$ *iterations of the CG solver in Algorithm 1, we have the following:*

$$
\|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \frac{\gamma\sigma}{4\sqrt{n}}\sqrt{\mu}\quad\textit{and}\quad \|\mathbf{v}\|_2\leq\frac{\gamma\sigma}{4}\mu\,,
$$

where $\psi=\left(\frac{9n}{\sqrt{1-\gamma}}+\sigma\sqrt{\frac{n}{1-\gamma}}+\sqrt{n}\right)$ *and* $\tilde{\mathbf{f}}^{(t)}=\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^t-\mathbf{Q}^{-1/2}\mathbf{p}$ *is the residual of the solver.*

Proof Combining Lemma 12 and the condition in eqn. (7), we have

$$\|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \zeta^t \psi \sqrt{2\mu}. \tag{60}$$

Now, $\|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \frac{\gamma\sigma}{4\sqrt{n}}\sqrt{\mu}$ holds if $\sqrt{2}\psi\,\zeta^t\sqrt{\mu} \leq \frac{\gamma\sigma}{4\sqrt{n}}\sqrt{\mu}$, which holds if $\left(\frac{1}{\zeta}\right)^t \geq \frac{4\sqrt{2\,n}\,\psi}{\gamma\sigma}$. The last inequality holds for our choice of t. Next, combining Lemma 4 and eqn. (60) we get

$$\|\mathbf{v}\|_2 \leq \sqrt{3n\mu}\,\|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \sqrt{6n}\,\zeta^t\psi\mu$$

Therefore, $\|\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4}$ holds if $\sqrt{6n}\psi\,\zeta^t\psi\mu \leq \frac{\gamma\sigma\mu}{4}$, which holds for our choice of t. Now, fixing γ, σ, and ζ, after $t = \mathcal{O}(\log n)$ iterations of Algorithm 1 the conclusions of the lemma hold. ∎

F.2 Determining Step-size, Bounding the Number of Iterations, and Proof of Theorem 1

Let $(\hat{\Delta\mathbf{x}}, \hat{\Delta\mathbf{y}}, \hat{\Delta\mathbf{s}})$ respectively satisfies eqns. (17), (18) and (14b). We rewrite the systen in the following alternative form

$$\mathbf{A}\hat{\Delta\mathbf{x}} = -\,\mathbf{r}_p\,, \tag{61a}$$

$$\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} + \hat{\Delta\mathbf{s}} = -\,\mathbf{r}_d\,, \tag{61b}$$

$$\mathbf{X}\hat{\Delta\mathbf{s}} + \mathbf{S}\hat{\Delta\mathbf{x}} = -\,\mathbf{X}\mathbf{S}\,\mathbf{1}_n + \sigma\mu\,\mathbf{1}_n - \mathbf{v}\,. \tag{61c}$$

Indeed, first we now show how to satisfy eqns. (17), (18) and (14b) from eqn. (61). Pre-multiplying both sides of eqn. (61c) by $\mathbf{A}\mathbf{S}^{-1}$ and noting that $\mathbf{D}^2 = \mathbf{X}\mathbf{S}^{-1}$, we get

$$\begin{aligned}
&\mathbf{A}\mathbf{D}^2\hat{\Delta\mathbf{s}} + \mathbf{A}\hat{\Delta\mathbf{x}} = -\mathbf{A}\mathbf{X}\mathbf{1}_n + \sigma\mu\mathbf{A}\mathbf{S}^{-1}\mathbf{1}_n - \mathbf{A}\mathbf{S}^{-1}\mathbf{v}\\
\Rightarrow\;&\mathbf{A}\mathbf{D}^2\hat{\Delta\mathbf{s}} = \mathbf{r}_p - \mathbf{A}\mathbf{x} + \sigma\mu\mathbf{A}\mathbf{S}^{-1}\mathbf{1}_n - \mathbf{A}\mathbf{S}^{-1}\mathbf{v}.
\end{aligned} \tag{62}$$

Eqn. (62) holds as $\mathbf{A}\mathbf{X}\mathbf{1}_n = \mathbf{A}\mathbf{x}$ and, from eqn. (61a), $\mathbf{A}\hat{\Delta\mathbf{x}} = -\mathbf{r}_p$. Next, pre-multiplying eqn. (61b) by $\mathbf{A}\mathbf{D}^2$, we get

$$\begin{aligned}
&\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} + \mathbf{A}\mathbf{D}^2\hat{\Delta\mathbf{s}} = -\mathbf{A}\mathbf{D}^2\mathbf{r}_d\\
\Rightarrow\;&\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\hat{\Delta\mathbf{y}} = -\mathbf{r}_p + \mathbf{A}\mathbf{x} - \sigma\mu\mathbf{A}\mathbf{S}^{-1}\mathbf{1}_n - \mathbf{A}\mathbf{D}^2\mathbf{r}_d + \mathbf{A}\mathbf{S}^{-1}\mathbf{v} = \mathbf{p} + \mathbf{A}\mathbf{S}^{-1}\mathbf{v}.
\end{aligned} \tag{63}$$

The first equality in eqn. (63) follows from eqn. (62) and the definition of $\mathbf{p}$ in eqn. (16). This establishes eqn. (18). Eqn. (14b) directly follows from eqn. (61b). Finally, we get eqn. (17) by pre-multiplying eqn. (61c) by $\mathbf{S}^{-1}$.

Next, we define each new point traversed by the algorithm as $(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha))$, where

$$(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha)) := (\mathbf{x}, \mathbf{y}, \mathbf{s}) + \alpha(\hat{\Delta\mathbf{x}}, \hat{\Delta\mathbf{y}}, \hat{\Delta\mathbf{s}}) \tag{64}$$

$$\mu(\alpha) := \mathbf{x}(\alpha)^T\mathbf{s}(\alpha)/n \tag{65}$$

$$\mathbf{r}(\alpha) := \mathbf{r}\left(\mathbf{x}(\alpha), \mathbf{s}(\alpha), \mathbf{y}(\alpha)\right). \tag{66}$$

The goal in this section is to bound the number of iterations required by Algorithm 2. Towards that end, we bound the magnitude of the step size α. First, we provide an upper bound on α, which allows us to show that each new point $(\mathbf{x}(\alpha), \mathbf{s}(\alpha), \mathbf{y}(\alpha))$ traversed by the algorithm stays within the neighborhood $\mathcal{N}(\gamma)$. Second, we provide a lower bound on α, which allows us to bound the number of iterations required. We use multiple lemmas from [38], which we reproduce here, without their proofs.

First, we provide an upper bound on α, ensuring that each new point $(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha))$ traversed by the algorithm stays within the neighborhood $\mathcal{N}(\gamma)$.

Lemma 14 (Lemma 3.5 of [38]) *Assume $(\hat{\Delta\mathbf{x}}, \hat{\Delta\mathbf{y}}, \hat{\Delta\mathbf{s}})$ satisfies eqns. (61) for some $\sigma > 0$, $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$ (for $\gamma \in (0, 1)$), and $\|\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4}$. Then, $(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha)) \in \mathcal{N}(\gamma)$ for every scalar α such that*

$$0 \leq \alpha \leq \min\left\{1, \frac{\gamma\sigma\mu}{4\|\hat{\Delta\mathbf{x}} \circ \hat{\Delta\mathbf{s}}\|_\infty}\right\}. \tag{67}$$

We now provide a lower bound on the values of $\bar{\alpha}$ and the corresponding $\mu(\bar{\alpha})$; see Algorithm 2.

Lemma 15 (Lemma 3.6 of [38]) *In each iteration of Algorithm 2, if* $\|\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4}$*, then the step size* $\bar{\alpha}$ *satisfies*

$$\bar{\alpha} \geq \min\left\{1, \frac{\min\{\gamma\sigma, (1-\frac{5}{4}\sigma)\}\mu}{4\|\hat{\Delta x} \circ \hat{\Delta s}\|_\infty}\right\} \tag{68}$$

and

$$\mu(\bar{\alpha}) = \Big[1 - \frac{\bar{\alpha}}{2}(1 - \frac{5}{4}\sigma)\Big]\mu. \tag{69}$$

At this point, we have provided a lower bound (eqn. (68)) for the allowed values of the step size $\bar{\alpha}$. Next, we show that this lower bound is bounded away from zero. From eqn. (68) this is equivalent to showing that $\|\hat{\Delta \mathbf{x}} \circ \hat{\Delta \mathbf{s}}\|_\infty$ is bounded.

Lemma 16 (Lemma 3.7 of [38] (slightly modified)) *Let* $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0)$ *be the initial point with* $(\mathbf{x}^0, \mathbf{s}^0) > 0$ *and* $(\mathbf{x}^0, \mathbf{s}^0) \geq (\mathbf{x}^*, \mathbf{s}^*)$ *for some* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*) \in \mathcal{S}$*. Let* $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$ *be such that* $\mathbf{r} = \eta\mathbf{r}^0$ *for some* $\eta \in [0,1]$ *and* $\|\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4}$*. Then, the search direction* $(\hat{\Delta \mathbf{x}}, \hat{\Delta \mathbf{y}}, \hat{\Delta \mathbf{s}})$ *produced by Algorithm 2 at each iteration satisfies*

$$\max\{\|\mathbf{D}^{-1}\hat{\Delta \mathbf{x}}\|_2, \|\mathbf{D}\hat{\Delta \mathbf{s}}\|_2\} \leq \left(1 + \frac{\sigma^2}{1-\gamma} - 2\sigma\right)^{1/2}\sqrt{n\mu} + \frac{6n}{\sqrt{(1-\gamma)}}\sqrt{\mu} + \frac{\gamma\sigma}{4\sqrt{1-\gamma}}\sqrt{\mu}. \tag{70}$$

We should note here that the above lemma is slightly different than Lemma 3.7 of [38]. Indeed, Lemma 3.7 of [38] actually proves the following bound:

$$\max\{\|\mathbf{D}^{-1}\hat{\Delta \mathbf{x}}\|_2, \|\mathbf{D}\hat{\Delta \mathbf{s}}\|_2\} \leq \left(1 + \frac{\sigma^2}{1-\gamma} - 2\sigma\right)^{1/2}\sqrt{n\mu} + \frac{6n}{\sqrt{(1-\gamma)}}\sqrt{\mu} + \frac{\gamma\sigma}{4\sqrt{n}}\sqrt{\mu}\,. \tag{71}$$

Notice that there is slight difference in the last term in the right-hand side, which does not asymptotically change the bound. The underlying reason for this difference is the fact that [38] constructed the vector $\mathbf{v}$ differently. In our case, we need to bound $\|(\mathbf{XS})^{-1/2}\mathbf{v}\|_2$, which we do as follows:

$$\|(\mathbf{XS})^{-1/2}\mathbf{v}\|_2 \leq \|(\mathbf{XS})^{-1/2}\|_2\,\|\mathbf{v}\|_2 \leq \frac{1}{\min_i\sqrt{x_i s_i}}\,\frac{\gamma\sigma\mu}{4}\,, \tag{72}$$

where in the above expression we use the fact that $\|(\mathbf{XS})^{-1/2}\|_2 = \frac{1}{\min_i\sqrt{x_i s_i}}$. Now as $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$, we further have $x_i s_i \geq (1-\gamma)\mu$ for all $i = 1 \ldots n$. Combining this with eqn. (72), we get

$$\|(\mathbf{XS})^{-1/2}\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4\sqrt{(1-\gamma)\mu}} = \frac{\gamma\sigma}{4\sqrt{1-\gamma}}\sqrt{\mu}. \tag{73}$$

On the other hand, [38] had a different construction of $\mathbf{v}$ for which $\|(\mathbf{XS})^{-1/2}\mathbf{v}\|_2 = \|\tilde{\mathbf{f}}^{(t)}\|_2$ holds. Therefore they had the following bound:

$$\|(\mathbf{XS})^{-1/2}\mathbf{v}\|_2 = \|\tilde{\mathbf{f}}^{(t)}\|_2 \leq \frac{\gamma\sigma}{4\sqrt{n}}\sqrt{\mu}.$$

Also, note that after correcting the approximation error of the CG solver using $\mathbf{v}$, the primal and dual residuals $\mathbf{r} = (\mathbf{r}_p, \mathbf{r}_d)$ corresponding to an iterate $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\gamma)$ always lie on the line segment between zero and $\mathbf{r}^{(0)}$. In other words, $\mathbf{r} = \eta\mathbf{r}^{(0)}$ always holds for some $\eta \in [0,1]$. This was formally proven in Lemma 3.3 of [38].

The next lemma bounds the number of iterations that Algorithm 2 needs when started with an infeasible point that is sufficiently positive.

Lemma 17 (Theorem 2.6 of [38]) *Assume that the constants* γ *and* σ *are such that* $\max\{\gamma^{-1}, (1-\gamma)^{-1}, \sigma^{-1}, (1-\frac{5}{4}\sigma)^{-1}\} = \mathcal{O}(1)$*. Let the initial point* $(\mathbf{x}^0, \mathbf{s}^0, \mathbf{y}^0)$ *satisfy* $(\mathbf{x}^0, \mathbf{s}^0) \geq (\mathbf{x}^*, \mathbf{s}^*)$ *for some* $(\mathbf{x}^*, \mathbf{s}^*, \mathbf{y}^*) \in \mathcal{S}$ *and* $\|\mathbf{v}\|_2 \leq \frac{\gamma\sigma\mu}{4}$*. Algorithm 2 generates an iterate* $(\mathbf{x}^k, \mathbf{s}^k, \mathbf{y}^k)$ *satisfying* $\mu_k \leq \epsilon\mu_0$ *and* $\|\mathbf{r}^k\|_2 \leq \epsilon\|\mathbf{r}^0\|_2$ *after* $\mathcal{O}(n^2 \log {}^1\!/\epsilon)$ *iterations.*

Finally, Theorem 1 follows from Lemmas 13 and 17.

Appendix G Additional Notes on Experiments

Problem	Size	Sketch IPM w/ Precond. CG				Stand. IPM w/ Unprec. CG			IPM w/ Dir.
	$(m \times N)$	w	In. It.	Out. It.	κ_{Sk}	In. It.	Out. It.	κ_{Stan}	Out. It.
ARCENE	$(100 \times 10K)$	200	**30**	50	38.09	**1.1K**	59	4.4×10^8	50
DEXTER	$(300 \times 20K)$	500	**39**	39	75.42	**4.6K**	39	7.6×10^9	39
DrivFace	$(606 \times 6.4K)$	$1K$	**50**	42	68.87	**139K**	43	17×10^{12}	42
Gene RNA	$(801 \times 20K)$	$2K$	**27**	44	20.03	**101K**	208	4.7×10^{12}	44

Table 1: Comparison of (our) sketched IPM with CG, standard IPM with CG, and Standard IPM with a direct solver, for the ℓ_1-SVM problem on UCI Machine Learning Repository [20] data sets. Across all, $\tau = 10^{-9}$ and a relative error of 10^{-3} or less was achieved. We define $\kappa_{\text{Sk}} = \kappa(\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\mathsf{T}\mathbf{Q}^{-1/2})$ and $\kappa_{\text{Stan}} = \kappa(\mathbf{A}\mathbf{D}^2\mathbf{A}^T)$.

G.1 Support Vector Machines (SVMs)

The classical ℓ_1-SVM problem is as follows. We consider the task of fitting an SVM to data pairs $S = \{(x_i, y_i)\}_{i=1}^m$, where $x_i \in \mathbb{R}^N$ and $y_i \in \{+1, -1\}$ is a label for each data pair. Here, m is the number of training points, and N is the feature dimension. The SVM problem with an ℓ_1 regularizer has the following form.

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \|w\|_1 \\ \text{subject to} \quad & y_i(w^T x_i + b') \geq 1, \quad \forall i \in [m]. \end{aligned} \tag{74}$$

This problem can be written as an LP by introducing the variables w^+ and w^-, where $w = w^+ - w^-$. The objective becomes $\sum_j^n w_j^+ + w_j^-$, and we constrain $w_i^+ \geq 0$ and $w_i^- \geq 0$. Note that the size of the constraint matrix in the LP becomes $(m \times (2N+1))$, where m is the number of training points, and N is the feature dimension.

G.2 Random Data

We generate random synthetic instances of linear programs as follows. To generate $A \in \mathbb{R}^{m \times n}$, we set $a_{ij} \sim_{i.i.d.} U(0,1)$ with probability p and $a_{ij} = 0$ otherwise. We then add min$\{m, n\}$ i.i.d. draws from $U(0,1)$ to the main diagonal, to ensure each row of A has at least one nonzero entry. We set $b = Ax + 0.1z$, where x and z are random vectors drawn from $N(0,1)$. Finally, we set $c \sim N(0,1)$.

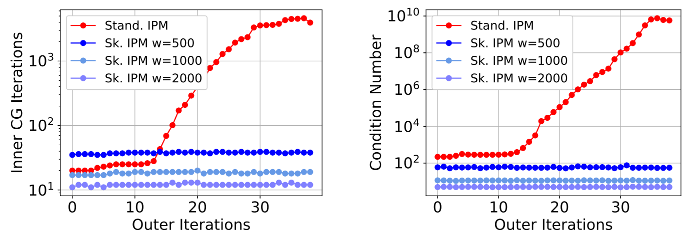
G.3 Real Data Descriptions

The following is how we made use a gene expression cancer RNA-Sequencing data set, taken from the UCI Machine Learning repository. It is part of the RNA-Seq (HiSeq) PANCAN data set [49], and is a random extraction of gene expressions from patients who have different types of tumors: BRCA, KIRC, COAD, LUAD and PRAD. We considered the binary classification task of identifying BRCA versus other types.

The following is how we made use of the DrivFace data set taken from the UCI Machine Learning repository. In the DrivFace data set, each sample corresponds to an image of a human subject, taken while driving in real scenarios. Each image is labeled as corresponding to one of 3 possible gaze directions (left, straight, or right). We considered the binary classification task of identifying two different gaze directions: (straight, or to either side left or right).

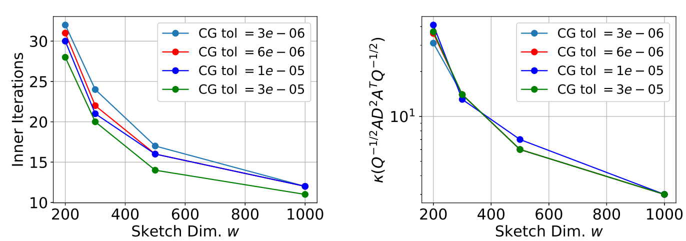
G.4 Additional Experiments

Here we include additional experiments. Figure 2 illustrates the convergence and conditioning behavior for the DEXTER data set. We see a similar behavior as found for the ARCENE data set in Figure 1. Figure 3 displays more results for the ARCENE data set.



(a) (b)

Figure 2: *DEXTER data set*: Our algorithm (Sk. IPM) requires an order of magnitude fewer inner iterations than the Standard IPM with CG, at each outer iteration, as demonstrated in (a). This is possible due to the improved conditioning of $\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\mathbf{Q}^{-1/2}$ compared to $\mathbf{A}\mathbf{D}^2\mathbf{A}^T$, demonstrated in (b). For all, *tolCG* $= 10^{-5}$, $\tau = 10^{-9}$.



(a) (b)

Figure 3: *ARCENE data set*: As w increases, (a) the number of inner iterations decreases, and is relatively robust to *tolCG*, and, (b) the condition number decreases as well.