

Appendix A Coordinate Wise Convergence vs Convergence in Objective Value

For an algorithm with a $\log \frac{1}{\varepsilon}$ dependence of the running time for computing a $(1 + \varepsilon)$ -approximate solution, like p -IRLS, the guarantee can be translated into a guarantee for convergence in the solution without any significant loss in the runtime complexity of the method. We demonstrate this theoretically and experimentally below.

Lemma A.1. *If \mathbf{x} is a $(1 + \delta)$ -approximate solution and \mathbf{x}^* is the optimum, then*

$$\|\mathbf{x} - \mathbf{x}^*\|_\infty \leq \frac{2m^{\frac{1}{2}}}{\sigma_{\min}(\mathbf{A})} \left(\frac{2\delta}{m} \right)^{\frac{1}{p}} \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p,$$

where $\sigma_{\min}(\mathbf{A})$ is the smallest singular value of \mathbf{A} .

Proof. Given that \mathbf{x} is a $(1 + \delta)$ -approximate solution, using Lemma B.1 we can write the following lower bound on the objective value:

$$(1 + \delta) \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p \geq \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p + p(\mathbf{A}\mathbf{x}^* - \mathbf{b})^\top \mathbf{R}\mathbf{A}(\mathbf{x} - \mathbf{x}^*) \\ + p/8 \mathbf{A}(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{A}^\top \mathbf{R}\mathbf{A}(\mathbf{x} - \mathbf{x}^*) + 2^{-(p+1)} \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}^*\|_p^p,$$

where $\mathbf{R} = \text{diag}(|\mathbf{A}\mathbf{x}^* - \mathbf{b}|^{p-2})$. Since the gradient at \mathbf{x}^* is 0, simplifying, we get, $2^{p+1}\delta \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p \geq \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}^*\|_p^p$. Now, translating between various norms, we obtain,

$$\|\mathbf{x} - \mathbf{x}^*\|_\infty \leq \frac{1}{\sigma_{\min}(\mathbf{A})} \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}^*\|_2 \leq \frac{m^{\frac{1}{2} - \frac{1}{p}}}{\sigma_{\min}(\mathbf{A})} \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}^*\|_p \leq \frac{2m^{\frac{1}{2}}}{\sigma_{\min}(\mathbf{A})} \left(\frac{2\delta}{m} \right)^{\frac{1}{p}} \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p.$$

□

We can achieve the guarantee $\|\mathbf{x} - \mathbf{x}^*\|_\infty \leq \varepsilon \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p$ by picking $\delta = \left(\frac{\varepsilon \sigma_{\min}(\mathbf{A})}{4m} \right)^p$. This gives $\log \frac{m}{\delta} = O(p \log \frac{m}{\sigma_{\min}(\mathbf{A})\varepsilon})$, and hence a total iteration count of $O(p^{4.5} m^{\frac{p-2}{2(p-1)}} \log \frac{m}{\sigma_{\min}(\mathbf{A})\varepsilon})$. Asymptotically, the running time bound is only off by a factor of p if we wish to measure the convergence in ℓ_∞ -norm, as long as $\log \frac{1}{\sigma_{\min}(\mathbf{A})} = O(\log \frac{m}{\varepsilon})$.

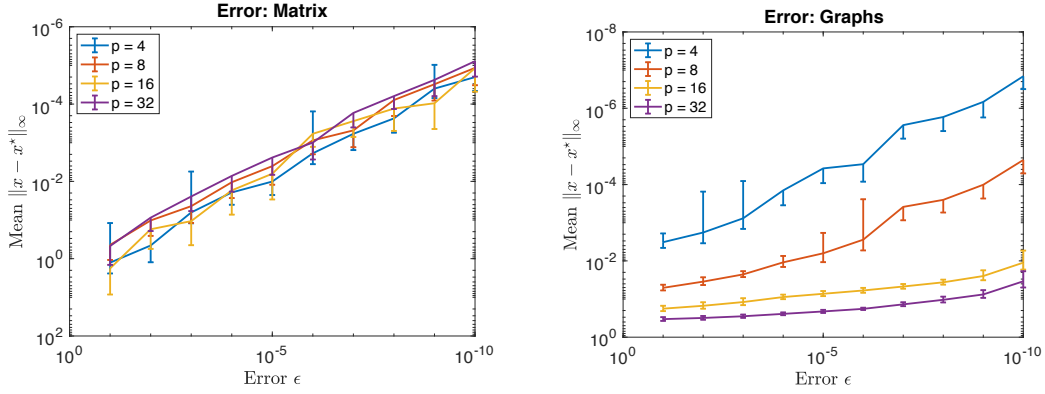


Figure 5: Maximum coordinate wise difference with the optimum vs accuracy to which the objective is close to the optimum, for both graphs and random matrix instances.

We also demonstrate this relation experimentally. The plots in Figure 5 demonstrate the average resulting ℓ_∞ norm deviation for the solution computed, as we change the ε parameter used in the algorithm. We use the instances described in the paper; matrices of size 1000×800 and graphs with 1000 nodes. For each instance, we: 1) find a very high accuracy solution, by choosing a very small $\varepsilon \sim 10^{-25}$, 2) scale the problem so that the optimum value is 1, and run the algorithm again to find the optimum solution \mathbf{x}^* . 3) Now we have a problem such that $\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p = 1$, we run the

algorithm again with various values of ε , to obtain solutions $\mathbf{x}(\varepsilon)$ and plot $\|\mathbf{x}(\varepsilon) - \mathbf{x}^*\|_\infty$ (averaged over 20 samples). These results are very much in agreement with the theoretical $\varepsilon^{\frac{1}{p}}$ dependence proved above. (Note that the error bars indicate $\log(\text{mean} \pm \text{std})$ so they are missing on one side when $\text{mean} < \text{std}$.)

Appendix B Proofs from Section 3

B.1 Proof of Lemma 3.2

Lemma 3.2. (*Iterative Refinement*). *Let $p \geq 2$, and $\kappa \geq 1$. Starting from $\mathbf{x}^{(0)} = \arg \min_{C_{x=d}} \|\mathbf{Ax} - \mathbf{b}\|_2^2$, and iterating as, $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \Delta$, where Δ is a κ -approximate solution to the residual problem (Definition 2.1), we get an ε -approximate solution to (3) in at most $O\left(p^2 \kappa \log\left(\frac{m}{\varepsilon}\right)\right)$ calls to a κ -approximate solver for the residual problem.*

We first show that we can upper and lower bound the change in objective by a linear term plus a quadratically smoothed function.

Lemma B.1. *For any \mathbf{x}, Δ and $p \geq 2$, we have for $\mathbf{r} = |\mathbf{x}|^{p-2}$ and $\mathbf{g} = p|\mathbf{x}|^{p-2}\mathbf{x}$,*

$$\frac{p}{8} \sum_e \mathbf{r}_e \Delta_e^2 + \frac{1}{2^{p+1}} \|\Delta\|_p^p \leq \|\mathbf{x} + \Delta\|_p^p - \|\mathbf{x}\|_p^p - \mathbf{g}^\top \Delta \leq 2p^2 \sum_e \mathbf{r}_e \Delta_e^2 + p^p \|\Delta\|_p^p.$$

The proof of the above lemma is long and hence deferred to the end of this section. Applying the above lemma on our objective we get,

$$\frac{p}{8} (\mathbf{A}\Delta)^\top \mathbf{R} \mathbf{A} \Delta + \frac{1}{2^{p+1}} \|\mathbf{A}\Delta\|_p^p \leq \|\mathbf{A}(\mathbf{x} + \Delta) - \mathbf{b}\|_p^p - \|\mathbf{Ax} - \mathbf{b}\|_p^p - \mathbf{g}^\top \mathbf{A} \Delta \leq 2p^2 (\mathbf{A}\Delta)^\top \mathbf{R} \mathbf{A} \Delta + p^p \|\mathbf{A}\Delta\|_p^p, \quad (5)$$

where \mathbf{R} is the diagonal matrix with entries $|\mathbf{Ax} - \mathbf{b}|^{p-2}$ and $\mathbf{g} = p\mathbf{R}(\mathbf{Ax} - \mathbf{b})$. We next show the relation between the residual problem defined in the preliminaries and the change in objective value when \mathbf{x} is updated by Δ .

Lemma B.2. *For any \mathbf{x}, Δ and $p \geq 2$ and $\lambda = 16p$,*

$$\gamma(\Delta) \leq \|\mathbf{Ax} - \mathbf{b}\|_p^p - \|\mathbf{A}(\mathbf{x} - \Delta) - \mathbf{b}\|_p^p,$$

and

$$\|\mathbf{Ax} - \mathbf{b}\|_p^p - \|\mathbf{A}(\mathbf{x} - \lambda\Delta) - \mathbf{b}\|_p^p \leq \lambda\gamma(\Delta).$$

Proof. The first inequality directly follows from (5). For the second inequality,

$$\begin{aligned} \|\mathbf{Ax} - \mathbf{b}\|_p^p - \|\mathbf{A}(\mathbf{x} - \lambda\Delta) - \mathbf{b}\|_p^p &\leq \lambda \mathbf{g}^\top \Delta - \lambda^2 \frac{p}{8} \Delta^\top \mathbf{A}^\top \mathbf{R} \mathbf{A} \Delta - \lambda^p \frac{1}{2^{p+1}} \|\mathbf{A}\Delta\|_p^p \\ &= \lambda \left(\mathbf{g}^\top \Delta - \lambda \frac{p}{8} \Delta^\top \mathbf{A}^\top \mathbf{R} \mathbf{A} \Delta - \lambda^{p-1} \frac{1}{2^{p+1}} \|\mathbf{A}\Delta\|_p^p \right) \\ &\leq \lambda \left(\mathbf{g}^\top \mathbf{A} \Delta - 2p^2 \Delta^\top \mathbf{A}^\top \mathbf{R} \mathbf{A} \Delta - p^p \|\mathbf{A}\Delta\|_p^p \right). \end{aligned}$$

□

B.1.1 Proof of Lemma Iterative Refinement

Proof. Let $\tilde{\Delta}$ be a κ -approximate solution to the residual problem. Using this fact and Lemma B.2 for $\Delta = \frac{\mathbf{x} - \mathbf{x}^*}{\lambda}$, we get,

$$\gamma(\tilde{\Delta}) \geq \frac{1}{\kappa} \gamma(\Delta^*) \geq \frac{1}{\kappa} \gamma\left(\frac{\mathbf{x} - \mathbf{x}^*}{\lambda}\right) \geq \frac{1}{\lambda \kappa} \left(\|\mathbf{Ax} - \mathbf{b}\|_p^p - OPT \right).$$

Also,

$$\begin{aligned}
\left\| \mathbf{A}(\mathbf{x} - \tilde{\Delta}) - \mathbf{b} \right\|_p^p - OPT &\leq \left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_p^p - \gamma(\tilde{\Delta}) - OPT \\
&\leq \left(\left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_p^p - OPT \right) - \frac{1}{\lambda\kappa} \left(\left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_p^p - OPT \right) \\
&= \left(1 - \frac{1}{\lambda\kappa} \right) \left(\left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_p^p - OPT \right).
\end{aligned}$$

Now, after t iterations,

$$\left\| \mathbf{A}(\mathbf{x}^{(t)} - \tilde{\Delta}) - \mathbf{b} \right\|_p^p - OPT \leq \left(1 - \frac{1}{\lambda\kappa} \right)^t \left(\left\| \mathbf{A}\mathbf{x}^{(0)} - \mathbf{b} \right\|_p^p - OPT \right) \leq \left(1 - \frac{1}{\lambda\kappa} \right)^t m^{(p-2)/2} OPT.$$

Thus, for our value of $\lambda = 16p, 8p^2\kappa \log(m/\varepsilon)$ iterations suffice to obtain a $(1 + \varepsilon)$ -approximate solution. \square

B.1.2 Proof of Lemma B.1

Proof. To show this, we show that the above holds for all coordinates. For a single coordinate, the above expression is equivalent to proving,

$$\frac{p}{8}|x|^{p-2}\Delta^2 + \frac{1}{2^{p+1}}|\Delta|^p \leq |x + \Delta|^p - |x|^p - p|x|^{p-1} \operatorname{sgn}(x)\Delta \leq 2p^2|x|^{p-2}\Delta^2 + p^p|\Delta|^p.$$

Let $\Delta = \alpha x$. Since the above clearly holds for $x = 0$, it remains to show for all α ,

$$\frac{p}{8}\alpha^2 + \frac{1}{2^{p+1}}|\alpha|^p \leq |1 + \alpha|^p - 1 - p\alpha \leq 2p^2\alpha^2 + p^p|\alpha|^p.$$

1. $\alpha \geq 1$:

In this case, $1 + \alpha \leq 2\alpha \leq p \cdot \alpha$. So, $|1 + \alpha|^p \leq p^p|\alpha|^p$ and the right inequality directly holds. To show the other side, let

$$h(\alpha) = (1 + \alpha)^p - 1 - p\alpha - \frac{p}{8}\alpha^2 - \frac{1}{2^{p+1}}\alpha^p.$$

We have,

$$h'(\alpha) = p(1 + \alpha)^{p-1} - p - \frac{p}{4}\alpha - \frac{p}{2^{p+1}}\alpha^{p-1}$$

and

$$h''(\alpha) = p(p-1)(1 + \alpha)^{p-2} - \frac{p}{4} - \frac{p(p-1)}{2^{p+1}}\alpha^{p-2} \geq 0.$$

Since $h''(\alpha) \geq 0$, $h'(\alpha) \geq h'(1) \geq 0$. So h is an increasing function in α and $h(\alpha) \geq h(1) \geq 0$.

2. $\alpha \leq -1$:

Now, $|1 + \alpha| \leq 1 + |\alpha| \leq p \cdot |\alpha|$, and $2\alpha^2p^2 - |\alpha|p \geq 0$. As a result,

$$|1 + \alpha|^p \leq -|\alpha|p + 2\alpha^2p^2 + p^p \cdot |\alpha|^p$$

which gives the right inequality. Consider,

$$h(\alpha) = |1 + \alpha|^p - 1 - p\alpha - \frac{p}{8}\alpha^2 - \frac{1}{2^{p+1}}|\alpha|^p.$$

$$h'(\alpha) = -p|1 + \alpha|^{p-1} - p - \frac{p}{4}\alpha + p\frac{1}{2^{p+1}}|\alpha|^{p-1}.$$

Let $\beta = -\alpha$. The above expression now becomes,

$$-p(\beta - 1)^{p-1} - p + \frac{p}{4}\beta + p\frac{1}{2^{p+1}}\beta^{p-1}.$$

We know that $\beta \geq 1$. When $\beta \geq 2$, $\frac{\beta}{2} \leq \beta - 1$ and $\frac{\beta}{2} \leq \left(\frac{\beta}{2}\right)^{p-1}$. This gives us,

$$\frac{p}{4}\beta + p\frac{1}{2^{p+1}}\beta^{p-1} \leq \frac{p}{2}\left(\frac{\beta}{2}\right)^{p-1} + \frac{p}{2}\left(\frac{\beta}{2}\right)^{p-1} \leq p(\beta - 1)^{p-1}$$

giving us $h'(\alpha) \leq 0$ for $\alpha \leq -2$. When $\beta \leq 2$, $\frac{\beta}{2} \geq \left(\frac{\beta}{2}\right)^{p-1}$ and $\frac{\beta}{2} \leq 1$.

$$\frac{p}{4}\beta + p\frac{1}{2^{p+1}}\beta^{p-1} \leq \frac{p}{2} \cdot \frac{\beta}{2} + \frac{p}{2} \cdot \frac{\beta}{2} \leq p$$

giving us $h'(\alpha) \leq 0$ for $-2 \leq \alpha \leq -1$. Therefore, $h'(\alpha) \leq 0$ giving us, $h(\alpha) \geq h(-1) \geq 0$, thus giving the left inequality.

3. $|\alpha| \leq 1$:

Let $s(\alpha) = 1 + p\alpha + 2p^2\alpha^2 + p^p|\alpha|^p - (1 + \alpha)^p$. Now,

$$s'(\alpha) = p + 4p^2\alpha + p^{p+1}|\alpha|^{p-1}\text{sgn}(\alpha) - p(1 + \alpha)^{p-1}.$$

When $\alpha \leq 0$, we have,

$$s'(\alpha) = p + 4p^2\alpha - p^{p+1}|\alpha|^{p-1} - p(1 + \alpha)^{p-1}.$$

and

$$s''(\alpha) = 4p^2 + p^{p+1}(p-1)|\alpha|^{p-2} - p(p-1)(1 + \alpha)^{p-1} \geq 2p^2 + p^{p+1}(p-1)|\alpha|^{p-2} - p(p-1) \geq 0.$$

So s' is an increasing function of α which gives us, $s'(\alpha) \leq s'(0) = 0$. Therefore s is a decreasing function, and the minimum is at 0 which is 0. This gives us our required inequality for $\alpha \leq 0$. When $\alpha \geq \frac{1}{p-1}$, $1 + \alpha \leq p \cdot \alpha$ and $s'(\alpha) \geq 0$. We are left with the range $0 \leq \alpha \leq \frac{1}{p-1}$. Again, we have,

$$\begin{aligned} s''(\alpha) &= 4p^2 + p^{p+1}(p-1)|\alpha|^{p-2} - p(p-1)(1 + \alpha)^{p-1} \\ &\geq 4p^2 + p^{p+1}(p-1)|\alpha|^{p-2} - p(p-1)\left(1 + \frac{1}{p-1}\right)^{p-1} \\ &\geq 4p^2 + p^{p+1}(p-1)|\alpha|^{p-2} - p(p-1)e, \text{ When } p \text{ gets large the last term approaches } e \\ &\geq 0. \end{aligned}$$

Therefore, s' is an increasing function, $s'(\alpha) \geq s'(0) = 0$. This implies s is an increasing function, giving, $s(\alpha) \geq s(0) = 0$ as required.

To show the other direction,

$$h(\alpha) = (1 + \alpha)^p - 1 - p\alpha - \frac{p}{8}\alpha^2 - \frac{1}{2^{p+1}}|\alpha|^p \geq (1 + \alpha)^p - 1 - p\alpha - \frac{p}{8}\alpha^2 - \frac{p}{8}\alpha^2 = (1 + \alpha)^p - 1 - p\alpha - \frac{p}{4}\alpha^2.$$

Now, since $p \geq 2$,

$$\begin{aligned} &\left((1 + \alpha)^{p-2} - 1\right)\text{sgn}(\alpha) \geq 0 \\ \Rightarrow &\left((1 + \alpha)^{p-1} - 1 - \alpha\right)\text{sgn}(\alpha) \geq 0 \\ \Rightarrow &\left(p(1 + \alpha)^{p-1} - p - \frac{p}{2}\alpha\right)\text{sgn}(\alpha) \geq 0 \end{aligned}$$

We thus have, $h'(\alpha) \geq 0$ when α is positive and $h'(\alpha) \leq 0$ when α is negative. The minimum of h is at 0 which is 0. This concludes the proof of this case. □

B.2 Proof of Lemma that Checks Progress in Objective

We will next prove the following Lemma which shows that we do not change i when we have the correct value of i .

Lemma B.3. (Check Progress). Let α_0 be as defined in line 4 of Algorithm 2 and $\tilde{\Delta}$ the solution of program 4. If $i/2 < \frac{(\|A\mathbf{x}^{(t)} - \mathbf{b}\|_p^p - \|A\mathbf{x}^* - \mathbf{b}\|_p^p)}{16p} \leq i$, then $\gamma(\alpha_0 \cdot \tilde{\Delta}) \geq \frac{\alpha_0 i}{4}$ and $(A\tilde{\Delta})^\top (R + sI)A\tilde{\Delta} \leq \lambda i/p^2$.

We require bounding the objective of program [4](#). To do that we first give a bound on a decision version of the residual problem, and then relate this problem with problem [4](#).

Lemma B.4. *Let i be such that the optimum of the residual problem, $\gamma(\Delta^*) \in (i/2, \lambda i]$. Then the following problem has optimum at most λi .*

$$\begin{aligned} \min_{\Delta \in \mathbb{R}^m} \quad & 2p^2(\mathbf{A}\Delta)^\top \mathbf{R}\mathbf{A}\Delta + p^p \|\mathbf{A}\Delta\|_p^p \\ & \mathbf{g}^\top \mathbf{A}\Delta = i/2 \\ & \mathbf{C}\Delta = 0. \end{aligned} \tag{6}$$

Proof. The assumption on the residual is

$$\gamma(\Delta^*) = \mathbf{g}^\top \mathbf{A}\Delta^* - 2p^2(\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* - p^p \|\mathbf{A}\Delta^*\|_p^p \in (i/2, \lambda i].$$

Since the last 2 terms are strictly non-positive, we must have, $\mathbf{g}^\top \mathbf{A}\Delta^* \geq i/2$. Since Δ^* is the optimum and satisfies $\mathbf{C}\Delta^* = 0$,

$$\frac{d}{d\lambda} \left(\mathbf{g}^\top \lambda \mathbf{A}\Delta^* - 2p^2 \lambda^2 (\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* - \lambda^p p^p \|\mathbf{A}\Delta^*\|_p^p \right)_{\lambda=1} = 0.$$

Thus,

$$\mathbf{g}^\top \mathbf{A}\Delta^* - 2p^2(\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* - p^p \|\mathbf{A}\Delta^*\|_p^p = 2p^2(\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* + (p-1)p^p \|\mathbf{A}\Delta^*\|_p^p.$$

Since $p \geq 2$, we get the following

$$2p^2(\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* + p^p \|\mathbf{A}\Delta^*\|_p^p \leq \mathbf{g}^\top \mathbf{A}\Delta^* - 2p^2(\mathbf{A}\Delta^*)^\top \mathbf{R}\mathbf{A}\Delta^* - p^p \|\mathbf{A}\Delta^*\|_p^p \leq \lambda i.$$

For notational convenience, let function $h_p(\mathbf{r}, \Delta) = 2p^2(\mathbf{A}\Delta)^\top \mathbf{R}\mathbf{A}\Delta + p^p \|\mathbf{A}\Delta\|_p^p$. Now, we know that, $\mathbf{g}^\top \mathbf{A}\Delta^* \geq i/2$ and $\mathbf{g}^\top \mathbf{A}\Delta^* - h_p(\mathbf{r}, \Delta^*) \leq \lambda i$. This gives,

$$i/2 \leq \mathbf{g}^\top \mathbf{A}\Delta^* \leq h_p(\mathbf{r}, \Delta^*) + \lambda i \leq 2\lambda i.$$

Let $\Delta = \delta \Delta^*$, where $\delta = \frac{i}{2\mathbf{g}^\top \mathbf{A}\Delta^*}$. Note that $\delta \in [1/4\lambda, 1]$. Now, $\mathbf{g}^\top \mathbf{A}\Delta = i/2$ and,

$$h_p(\mathbf{r}, \Delta) \leq \max\{\delta^2, \delta^p\} h_p(\mathbf{r}, \Delta^*) \leq \lambda i.$$

Note that this Δ satisfies the constraints of program [6](#) and has an optimum at most λi . So the optimum of the program must have an objective at most λi . \square

Claim B.5. *If the optimal objective of program [6](#) is at most Z , then the optimum objective of program [4](#) is at most $\frac{Z}{2p^2} + \frac{i^{(p-2)/p} Z^{2/p}}{2p^2}$.*

Proof. Let Δ^* denote the optimizer of [6](#) and $\tilde{\Delta}$ be the optimizer of [4](#). Since the optimum objective of [6](#) is at most Z , we have we have $\|\mathbf{A}\Delta^*\|_p^p \leq \frac{Z}{p^p}$. This implies that $\Delta^{*\top} \mathbf{A}^\top \mathbf{A}\Delta^* \leq \frac{Z^{2/p}}{p^2} m^{(p-2)/p}$. Since Δ^* is a feasible solution of [4](#), we have for our value of s ,

$$\tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s^{(t)} \mathbf{I}) \mathbf{A}\tilde{\Delta} \leq \Delta^{*\top} \mathbf{A}^\top (\mathbf{R} + s^{(t)} \mathbf{I}) \mathbf{A}\Delta^* \leq \frac{Z}{2p^2} + \frac{i^{(p-2)/p} Z^{2/p}}{2p^2}.$$

\square

Proof of Lemma [B.3](#)

Proof. Since,

$$i/2 < \frac{(\|\mathbf{A}\mathbf{x}^{(t)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p)}{16p} \leq i,$$

we know that the optimum of the residual problem lies between $(i/2, \lambda i]$ (Lemma B.2). From Lemma B.4 the optimum of the problem (6) is at most λi . Now, from Claim B.5, we know that $(\tilde{\mathbf{A}}\tilde{\Delta})^\top(\mathbf{R} + s\mathbf{I})(\tilde{\mathbf{A}}\tilde{\Delta}) \leq \lambda i/p^2$. Also, note that $\alpha_0 + k\alpha_0^{p-1} \leq \frac{1}{8\lambda}$. Consider the following,

$$\begin{aligned}
\gamma(\alpha_0 \cdot \tilde{\Delta}) &= \alpha_0 \mathbf{g}^\top \tilde{\mathbf{A}}\tilde{\Delta} - \alpha_0^2 2p^2 (\tilde{\mathbf{A}}\tilde{\Delta})^\top \mathbf{R}(\tilde{\mathbf{A}}\tilde{\Delta}) - \alpha_0^p p^p \|\tilde{\mathbf{A}}\tilde{\Delta}\|_p^p \\
&\geq \alpha_0 \mathbf{g}^\top \tilde{\mathbf{A}}\tilde{\Delta} - (\alpha_0^2 + k\alpha_0^p) 2p^2 (\tilde{\mathbf{A}}\tilde{\Delta})^\top (\mathbf{R} + s\mathbf{I})(\tilde{\mathbf{A}}\tilde{\Delta}) \\
&= \alpha_0 \left(\mathbf{g}^\top \tilde{\mathbf{A}}\tilde{\Delta} - (\alpha_0 + k\alpha_0^{p-1}) 2p^2 (\tilde{\mathbf{A}}\tilde{\Delta})^\top (\mathbf{R} + s\mathbf{I})(\tilde{\mathbf{A}}\tilde{\Delta}) \right) \\
&\geq \alpha_0 \left(\mathbf{g}^\top \tilde{\mathbf{A}}\tilde{\Delta} - \frac{1}{8\lambda} 2p^2 (\tilde{\mathbf{A}}\tilde{\Delta})^\top (\mathbf{R} + s\mathbf{I})(\tilde{\mathbf{A}}\tilde{\Delta}) \right) \\
&\geq \alpha_0 \left(\frac{i}{2} - \frac{1}{8\lambda} 2\lambda i \right) \\
&\geq \frac{\alpha_0}{4} i
\end{aligned}$$

□

B.3 Proof of Lemma 3.4

Lemma 3.4. (Invariant) *At every iteration of the while loop, we have $\mathbf{C}\mathbf{x}^{(t)} = \mathbf{d}$, $\frac{(\|\mathbf{A}\mathbf{x}^{(t)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p)}{16p} \leq i$ and $i \geq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_p^p m^{-(p-2)/2}$.*

Proof. We use induction to show this. Initially we set, $i = \|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_p^p / 16p$. When the optimum is not 0, this is greater than $(\|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p) / 16p$. When the optimum is 0, the initial solution (2-norm minimizer) will also give zero and we can stop our procedure. Therefore, the claim holds for $t = 1$. Suppose at iteration t the claim holds. Since the objective is non-increasing, we know that,

$$\|\mathbf{A}\mathbf{x}^{(t)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p \geq \|\mathbf{A}\mathbf{x}^{(t+1)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p.$$

Let $\tilde{\Delta}$ denote the solution returned in iteration $t + 1$. At iteration $t + 1$, if $(\|\mathbf{A}\mathbf{x}^{(t+1)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p) / 16p \in (i/2, i]$, from Lemma B.3, we will always have $\gamma(\alpha_0 \tilde{\Delta}) \geq \alpha_0 i / 4$ and $(\tilde{\mathbf{A}}\tilde{\Delta})^\top(\mathbf{R} + s\mathbf{I})\tilde{\mathbf{A}}\tilde{\Delta} \leq \lambda i / p^2$. So the algorithm does not reduce i and as a result our claim holds for $t + 1$. Otherwise, we know that $(\|\mathbf{A}\mathbf{x}^{(t+1)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p) / 16p \leq i/2$ and the algorithm might reduce i by half if either of the two conditions are true. However, the claim still holds. Therefore, i is always at least $(\|\mathbf{A}\mathbf{x}^{(t+1)} - \mathbf{b}\|_p^p - \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p) / 16p$.

We start with a solution $\mathbf{x}^{(0)}$ that minimizes the ℓ_2 norm. Therefore, the following holds,

$$\|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_p^p \leq \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p m^{(p-2)/2}.$$

The value of i is the minimum at termination. Therefore, it is sufficient to prove the above bound for the termination condition. Our condition gives us the following at termination (the left inequality holds because otherwise, we would have terminated in the previous iteration).

$$i \geq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p^p \geq \frac{i}{2}.$$

This implies,

$$i \geq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_p^p \geq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_p^p m^{-(p-2)/2}.$$

We next prove the second claim. Initially our solution satisfies $\mathbf{C}\mathbf{x}^{(0)} = \mathbf{d}$. Assuming the condition holds at iteration t , we show it for $t + 1$. At every iteration we solve for $\tilde{\Delta}$ under the constraint $\mathbf{C}\tilde{\Delta} = 0$. Our update rule, $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha \tilde{\Delta}$ gives us,

$$\mathbf{C}\mathbf{x}^{(t+1)} = \mathbf{C}\mathbf{x}^{(t)} - \alpha \mathbf{C}\tilde{\Delta} = \mathbf{d} - \alpha \cdot 0 = \mathbf{d}.$$

□

B.4 Proof of Lemma 3.3

Our approximation depends on the quantity α_0 which is defined in the algorithm. This depends on the value of k , the ratio of the p -norm term to the square term. Therefore, in order to bound the approximation, we first give a bound on k .

Lemma B.6. *Let $\tilde{\Delta}$ the optimum of (4) and let $k = \frac{p^p \|\mathbf{A}\tilde{\Delta}\|_p^p}{2p^2 \tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \tilde{\Delta}}$. If the optimum of (4) is at most $\lambda i / p^2$, then k is at most $(32pm)^{(p-2)/2}$ for $\lambda = 16p$. Let $\alpha_0 = \min \left\{ \frac{1}{16\lambda}, \frac{1}{(16\lambda k)^{1/(p-1)}} \right\}$. Then $\alpha_0 \geq \Omega(p^{-1/2} m^{-\frac{(p-2)}{2(p-1)}})$ when $p \leq m$.*

Proof. Since, $s\mathbf{I} \preceq \mathbf{R} + s\mathbf{I}$,

$$\|\mathbf{A}\tilde{\Delta}\|_2^2 = \tilde{\Delta}^\top \mathbf{A}^\top \mathbf{A} \tilde{\Delta} \leq \frac{1}{s} \tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \tilde{\Delta}$$

and,

$$\|\mathbf{A}\tilde{\Delta}\|_p^p \leq \|\mathbf{A}\tilde{\Delta}\|_2^p \leq \frac{1}{s} \left(\tilde{\Delta}^\top \mathbf{A}^\top \mathbf{A} \tilde{\Delta} \right)^{(p-2)/2} \tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \tilde{\Delta}.$$

We also have, $\tilde{\Delta}^\top \mathbf{A}^\top \mathbf{A} \tilde{\Delta} \leq \frac{2\lambda}{p^2} m^{(p-2)/p} i^{2/p}$. Combining these,

$$\begin{aligned} \frac{p^p \|\mathbf{A}\tilde{\Delta}\|_p^p}{2p^2 \tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \tilde{\Delta}} &\leq \frac{p^p}{2p^2} \frac{2m^{(p-2)/p}}{i^{(p-2)/p}} \left(\frac{2\lambda}{p^2} m^{(p-2)/p} i^{2/p} \right)^{(p-2)/2} \\ &\leq (4\sqrt{2})^{p-2} p^{(p-2)/2} m^{(p-2)/p} m^{(p-2)^2/2p} \\ &= (32pm)^{(p-2)/2} \end{aligned}$$

We can now find a bound on α_0 .

$$\begin{aligned} \alpha_0 &\geq \min \left\{ \Omega\left(\frac{1}{p}\right), \Omega\left(\frac{1}{p^{1/(p-1)} (pm)^{(p-2)/2(p-1)}}\right) \right\} \\ &\geq \min \left\{ \Omega\left(\frac{1}{p}\right), \Omega\left(\frac{1}{p^{1/2} m^{(p-2)/2(p-1)}}\right) \right\} \\ &\geq \Omega\left(\frac{1}{p^{1/2} m^{(p-2)/2(p-1)}}\right), \text{ assuming } p \leq m. \end{aligned}$$

□

Lemma 3.3. (Approximation). *Let $\mathbf{R}, \mathbf{g}, s, \alpha$ be as defined in lines (5), (6), (7) and (9) of Algorithm 1. Let α_0 be as defined in line (4) of Algorithm 2 and $\tilde{\Delta}$ be the solution of the following program,*

$$\arg \min_{\Delta} \Delta^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \Delta \quad \text{s.t.} \quad \mathbf{g}^\top \mathbf{A} \Delta = i/2, \mathbf{C} \Delta = 0. \quad (4)$$

If $\tilde{\Delta}^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \tilde{\Delta} \leq \lambda i / p^2$ and $\gamma(\alpha_0, \tilde{\Delta}) \geq \frac{\alpha_0 i}{4}$, then $\alpha \cdot \tilde{\Delta}$ is an $O\left(p^{1.5} m^{\frac{p-2}{2(p-1)}}\right)$ -approximate solution to the residual problem.

Proof. In the algorithm we choose α such that given $\tilde{\Delta}$, $\alpha = \arg \min_{\delta} \|\mathbf{A}(\mathbf{x} - \delta \tilde{\Delta}) - \mathbf{b}\|_p^p$. From our assumption, we also know that $\gamma(\alpha_0, \tilde{\Delta}) \geq \frac{\alpha_0 i}{4}$. Now, since the residual function is a convex function

with value zero at the zero vector, we know that $\gamma(\tilde{\Delta}/\lambda) \leq \frac{1}{\lambda}\gamma(\tilde{\Delta})$ (for our value of $\lambda = 16p$).

$$\begin{aligned}
\gamma(\alpha\tilde{\Delta}) &\geq \lambda\gamma(\alpha\tilde{\Delta}/\lambda) \\
&\geq \|\mathbf{Ax} - \mathbf{b}\|_p^p - \|\mathbf{A}(\mathbf{x} - \alpha\tilde{\Delta}) - \mathbf{b}\|_p^p \\
&\geq \|\mathbf{Ax} - \mathbf{b}\|_p^p - \|\mathbf{A}(\mathbf{x} - \alpha_0\tilde{\Delta}) - \mathbf{b}\|_p^p \\
&\geq \gamma(\alpha_0 \cdot \tilde{\Delta}) \\
&\geq \frac{\alpha_0}{4}i \\
&\geq \frac{\alpha_0}{4\lambda}OPT.
\end{aligned}$$

The last inequality follows from the fact that the optimum of the residual problem is at most λi . This is because, Lemma 3.4 shows that, $(\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p - \|\mathbf{Ax}^* - \mathbf{b}\|_p^p)/16p < i$. Now from Lemma B.2 we can conclude that the residual problem has optimum at most λi . Since we have from our assumption that the objective of (4) is at most $\lambda i/p^2$, from Lemma B.6 we can bound the factor $O(\lambda/\alpha_0) \leq O(p^{1+\frac{p-2}{2(p-1)}} m^{\frac{p-2}{2(p-1)}}) \leq O(p^{1.5} m^{\frac{p-2}{2(p-1)}})$.

□

B.5 Proof of Lemma 3.5

Lemma 3.5. (Termination). *Let i be such that $(\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p - \|\mathbf{Ax}^* - \mathbf{b}\|_p^p)/16p \in (i/2, i]$. Then,*

$$i \leq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p \Rightarrow \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p \leq (1+\varepsilon)OPT.$$

and,

$$\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p \leq (1+\varepsilon)OPT \Rightarrow i \leq 2\frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p.$$

Proof. We first show the forward implication. From the assumptions, we have,

$$\begin{aligned}
\frac{\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p - OPT}{16p} &\leq i \leq \frac{\varepsilon}{16p(1+\varepsilon)} \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p \\
\Rightarrow \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p \frac{1}{1+\varepsilon} &\leq OPT \\
\Rightarrow \|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p &\leq (1+\varepsilon)OPT.
\end{aligned}$$

For the other direction we have,

$$\frac{\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p}{1+\varepsilon} \leq OPT.$$

Thus,

$$i \leq 2\frac{\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p - OPT}{16p} \leq 2\frac{\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p}{16p} \left(1 - \frac{1}{1+\varepsilon}\right) \leq \frac{2\varepsilon}{1+\varepsilon} \frac{\|\mathbf{Ax}^{(t)} - \mathbf{b}\|_p^p}{16p}.$$

□

Appendix C Converting ℓ_p -Laplacian Minimization to Regression Form

Define the following terms:

- n denote the number of vertices.
- l denote the number of labels.
- \mathbf{B} denote the edge-vertex adjacency matrix.

- \mathbf{g} denote the vector of labels for the l labelled vertices.
- \mathbf{W} denote the diagonal matrix with weights of the edges.

Set $\mathbf{A} = \mathbf{W}^{1/p} \mathbf{B}$ and $\mathbf{b} = -\mathbf{B}[:, n : n + l] \mathbf{g}$. Now $\|\mathbf{Ax} - \mathbf{b}\|_p^p$ is equal to the ℓ_p laplacian and we can use our IRLS algorithm to find the \mathbf{x} that minimizes this.

Appendix D Solving ℓ_2 Problems under Subspace Constraints

D.1 Finding the Initial Solution

We want to solve:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{Ax} - \mathbf{b}\|_2^2 \\ & \mathbf{Cx} = \mathbf{d}. \end{aligned}$$

Using Lagrangian duality and noting that strong duality holds, we can write the above as,

$$\begin{aligned} L(\mathbf{x}, \mathbf{v}) &= \min_{\mathbf{x}} \max_{\mathbf{v}} (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) + \mathbf{v}^\top (\mathbf{d} - \mathbf{Cx}) \\ &= \max_{\mathbf{v}} \min_{\mathbf{x}} (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) + \mathbf{v}^\top (\mathbf{d} - \mathbf{Cx}). \end{aligned}$$

We first find \mathbf{x}^* that minimizes the above objective by setting the gradient with respect to \mathbf{x} to 0. We thus have,

$$\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \left(\frac{2\mathbf{A}^\top \mathbf{b} + \mathbf{C}^\top \mathbf{v}}{2} \right).$$

Using this value of \mathbf{x} we arrive at the following dual program.

$$L(\mathbf{v}) = \max_{\mathbf{v}} -\frac{1}{4} \mathbf{v}^\top \mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{C}^\top \mathbf{v} - \mathbf{b}^\top \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} - \mathbf{v}^\top \mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b} + \mathbf{v}^\top \mathbf{d},$$

which is optimized at,

$$\mathbf{v}^* = 2 \left(\mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{C}^\top \right)^{-1} \left(\mathbf{d} - \mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \right).$$

Strong duality also implies that $L(\mathbf{x}, \mathbf{v}^*)$ is optimized at \mathbf{x}^* , which gives us,

$$\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \left(\mathbf{A}^\top \mathbf{b} + \mathbf{C}^\top \left(\mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{C}^\top \right)^{-1} \left(\mathbf{d} - \mathbf{C} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \right) \right).$$

D.2 Solving [4](#)

At every iteration of the algorithm, we want to solve the following problem,

$$\begin{aligned} \min_{\Delta} \quad & \Delta^\top \mathbf{A}^\top (\mathbf{R} + s\mathbf{I}) \mathbf{A} \Delta \\ & \mathbf{g}^\top \mathbf{A} \Delta = i/2 \\ & \mathbf{C} \Delta = 0. \end{aligned}$$

The constraints can be combined and rewritten as, $\mathbf{C}' \Delta = \mathbf{d}'$ where,

$$\mathbf{C}' = \begin{bmatrix} \mathbf{C} \\ \mathbf{g}^\top \mathbf{A} \end{bmatrix}, \quad \mathbf{d}' = \begin{bmatrix} 0 \\ i/2 \end{bmatrix}.$$

Let $\mathbf{R}' = \mathbf{R} + s\mathbf{I}$. We now want to solve,

$$\begin{aligned} \min_{\Delta} \quad & \|\mathbf{R}'^{1/2} \mathbf{A} \Delta\|_2^2 \\ & \mathbf{C}' \Delta = \mathbf{d}'. \end{aligned}$$

Using a procedure similar as in the previous section, we get,

$$\Delta^* = (\mathbf{A}^\top \mathbf{R}' \mathbf{A})^{-1} \mathbf{C}'^\top \left(\mathbf{C}' (\mathbf{A}^\top \mathbf{R}' \mathbf{A})^{-1} \mathbf{C}'^\top \right)^{-1} \mathbf{d}'$$