Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [N/A]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 2
   (b) Did you include complete proofs of all theoretical results? [Yes] See Appendix for complete proofs
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Algorithm for General-Self-Concordant Functions

In this section we will show how to use our algorithms for the following classes of general-self-concordant functions.

1. $6 > \nu \geq 2$: $\boldsymbol{f}$ is (N, ν)-g.s.c. and L-smooth.
2. $\nu < 2$: $\boldsymbol{f}$ is (N, ν)-g.s.c., L-smooth and μ-strongly convex.

We will use the following result to reduce these problems to $(M, 2)$-g.s.c. problems and use our algorithms.

Lemma A.1 (Prop 4. [STD19]). *Let $\boldsymbol{f}$ be (M, ν)-g.s.c. with $\nu > 0$. Then:*

(a) If $\nu \in (0, 3]$ and $\boldsymbol{f}$ is also strongly convex with strong convexity parameter $\mu > 0$ in ℓ_2-norm, then $\boldsymbol{f}$ is also $\left(\frac{M}{\sqrt{\mu^{3-\nu}}}, 3\right)$-g.s.c.

(b) If $\nu \geq 2$ and $\nabla \boldsymbol{f}$ is Lipschitz continuous with finite Lipschitz constant L in ℓ_2-norm, then $\boldsymbol{f}$ is also $\left(ML^{\frac{\nu}{2}-1}, 2\right)$-g.s.c.

We thus have the following result.

Theorem A.2. *For $\delta > 0$, $\boldsymbol{f}$ (N, ν)-g.s.c. $6 > \nu \geq 2$ and L-smooth, let $\overline{\boldsymbol{x}}$ be the solution returned by Algorithm 1 (with $\epsilon = 1$) applied to $\boldsymbol{f}(\boldsymbol{x})$. Now, Algorithm 2 with starting solution $\boldsymbol{x}^{(0)} = \overline{\boldsymbol{x}}$, applied to $\boldsymbol{f}$ finds $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and $\sum_i \boldsymbol{f}(\boldsymbol{P}\widetilde{\boldsymbol{x}}_i) \leq \sum_i \boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}_i^\star) + \delta$ in at most*

$$O\left(m^{1/3} N L^{\frac{\nu-2}{2}} R \log\left(\frac{\boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{f}(\boldsymbol{x}^\star))}{\delta}\right)\right)$$

calls to a linear system solver.

Proof. From Lemma A.1, $\boldsymbol{f}$ is $(NL^{(\nu-2)/2}, 2)$-g.s.c. We now use Lemma 3.3 with $M = NL^{(\nu-2)/2}$ followed by Theorem 4.6. □

Theorem A.3. *For $\delta > 0$, $\boldsymbol{f}$ (N, ν)-g.s.c. $2 > \nu \geq 0$ and L-smooth μ-strongly convex, let $\overline{\boldsymbol{x}}$ be the solution returned by Algorithm 1 (with $\epsilon = 1$) applied to $\boldsymbol{f}(\boldsymbol{x})$. Now, Algorithm 2 with starting solution $\boldsymbol{x}^{(0)} = \overline{\boldsymbol{x}}$, applied to $\boldsymbol{f}$ finds $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and $\sum_i \boldsymbol{f}(\boldsymbol{P}\widetilde{\boldsymbol{x}}_i) \leq \sum_i \boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}_i^\star) + \delta$ in at most*

$$O\left(m^{1/3} N \mu^{-\frac{3-\nu}{2}} L^{1/2} R \log\left(\frac{\boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{f}(\boldsymbol{x}^\star))}{\delta}\right)\right)$$

calls to a linear system solver.

Proof. From Lemma A.1, $\boldsymbol{f}$ is $(N\mu^{-\frac{3-\nu}{2}} L^{1/2}, 2)$-g.s.c. We now use Lemma 3.3with $M = N\mu^{-\frac{3-\nu}{2}} L^{1/2}$ followed by Theorem 4.6. □

B Missing Proofs

B.1 Proofs from Section 2

Definition B.1. *[Hessian Stability] For distance $r \in \mathbb{R}_{\geq 0}$ and function $\boldsymbol{d} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ acting on r, a function $\boldsymbol{f}$ is $(r, \boldsymbol{d}(r))$-hessian stable w.r.t. a norm $\|\cdot\|$ if for all $\boldsymbol{x}, \boldsymbol{y}$ such that $\|\boldsymbol{x} - \boldsymbol{y}\| \leq r$,*

$$\frac{1}{\boldsymbol{d}(r)} \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{y}) \preceq \boldsymbol{d}(r) \nabla^2 \boldsymbol{f}(\boldsymbol{x})$$

Lemma B.2 (Lemma 11 [CJJ+20]). *If $\boldsymbol{f}$ is a univariate M-quasi-self-concordant (q.s.c.) function, then $\boldsymbol{f}(\boldsymbol{x}) = \sum_i \boldsymbol{f}(\boldsymbol{x}_i)$ is (r, e^{Mr}) hessian stable in the ℓ_∞-norm.*

Lemma 2.4. *For $\epsilon > 0$, resistances $\boldsymbol{r}$ (Definition 2.3), with corresponding weights $\boldsymbol{w}$, we have*

$$\Psi(\boldsymbol{r}) \leq (1+\epsilon)\Phi(\boldsymbol{w}).$$

In addition, letting $\|\boldsymbol{P}\|_{\min} = \min_{\boldsymbol{Ax}=\boldsymbol{b}} \|\boldsymbol{Px}\|_2$ and $\|\boldsymbol{A}\|$ denote the operator norm of $\boldsymbol{A}$, we have

$$\Psi(\boldsymbol{r}) \geq \frac{\epsilon\Phi(\boldsymbol{w})}{mR^2} \frac{\|\boldsymbol{P}\|_{\min}^2 \|\boldsymbol{b}\|_2^2}{\|\boldsymbol{A}\|^2} \stackrel{\text{def}}{=} \Phi(\boldsymbol{w})L.$$

Proof. Let $\widetilde{\Delta}$ be the minimizer of $\Psi(\boldsymbol{r})$ and $\boldsymbol{x}^\star$ be the optimum of (1).

$$\begin{aligned}
\Psi(\boldsymbol{r}) &= \sum_i \boldsymbol{r}_i (\boldsymbol{P}\widetilde{\Delta})_i^2 \leq \sum_i \boldsymbol{r}_i (\boldsymbol{Px}^\star)_i^2 \\
&= \sum_i \left(\boldsymbol{f}''(\boldsymbol{w}_i) + \frac{\epsilon\Phi(\boldsymbol{w})}{m}\right) \frac{(\boldsymbol{Px}^\star)_i^2}{R^2} \\
&\leq \sum_i \boldsymbol{f}''(\boldsymbol{w}_i) + \frac{\epsilon\Phi(\boldsymbol{w})}{m} \cdot m, && \text{Since } \|\boldsymbol{Px}^\star\|_\infty \leq R \\
&= \Phi(\boldsymbol{w})(1+\epsilon)
\end{aligned}$$

We next look at a lower bound for Ψ. We note that, any solution to the oracle must satisfy $\boldsymbol{A}\widetilde{\Delta} = \boldsymbol{b}$. This implies, $\|\boldsymbol{A}\|\|\widetilde{\Delta}\|_2 \geq \|\boldsymbol{b}\|_2$, where $\|\cdot\|$ denotes the operator norm. Now,

$$\Psi(\boldsymbol{r}) \geq \frac{\epsilon\Phi(\boldsymbol{w})}{mR^2}\|\boldsymbol{P}\widetilde{\Delta}\|_2^2 \geq \frac{\epsilon\Phi(\boldsymbol{w})}{mR^2}\|\boldsymbol{P}\|_{\min}^2\|\widetilde{\Delta}\|_2^2 \geq \frac{\epsilon\Phi(\boldsymbol{w})}{mR^2}\frac{\|\boldsymbol{P}\|_{\min}^2\|\boldsymbol{b}\|_2^2}{\|\boldsymbol{A}\|^2}.$$

□

Lemma B.3.

$$\sum_i \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \leq (1+\epsilon)R\Phi(\boldsymbol{w})$$

Proof.

$$\begin{aligned}
\sum_i \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i &\leq \sqrt{\sum_i \boldsymbol{f}''(\boldsymbol{w}_i) \sum_i \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i^2} && \text{Cauchy Schwarz} \\
&\leq \sqrt{\Phi(\boldsymbol{w})}\sqrt{R^2\Psi(\boldsymbol{r})} \\
&\leq R\sqrt{\Phi(\boldsymbol{w})}\sqrt{(1+\epsilon)\Phi(\boldsymbol{w})} && \text{From Lemma 2.4} \\
&= R(1+\epsilon)\Phi(\boldsymbol{w}).
\end{aligned}$$

□

B.2 Proofs from Section 3

Change in Ψ

Lemma 3.1. *Let Ψ be as defined in 2.3. After t flow steps and k width reduction steps, we have,*

$$\Psi(\boldsymbol{r}^{(t,k)}) \geq \Psi(\boldsymbol{r}^{(0,0)})\left(1 + \frac{\epsilon^2\tau^2}{(1+\epsilon)^2 m}\right)^k \qquad \text{if } \boldsymbol{f}'' \text{ non-decreasing in } \boldsymbol{w},$$

$$\Psi(\boldsymbol{r}^{(t,k)}) \leq \Psi(\boldsymbol{r}^{(0,0)})\left(1 - \frac{\epsilon^2\tau^2}{2(1+\epsilon)^2 m}\right)^k \qquad \text{if } \boldsymbol{f}'' \text{ non-increasing in } \boldsymbol{w}.$$

Proof. We show this by induction. It is clear that this holds for $t = k = 0$. We know from Lemma C.2, for $\boldsymbol{r}' \geq \boldsymbol{r}$,

$$\Psi(\boldsymbol{r}') \geq \Psi(\boldsymbol{r}) + \sum_i \left(1 - \frac{\boldsymbol{r}_i}{\boldsymbol{r}'_i}\right)\boldsymbol{r}_i(\boldsymbol{P}\widetilde{\Delta})_i^2.$$

Since the weights are only increasing, this corresponds to the case $\boldsymbol{f}''$ is an increasing function. Similarly, when $\boldsymbol{f}''$ is a non-increasing function, we have the following bound: for $\boldsymbol{r}' \leq \boldsymbol{r}$ from Lemma C.1,

$$\Psi(\boldsymbol{r}') \leq \Psi(\boldsymbol{r}) - \frac{1}{2}\sum_i \left(1 - \frac{\boldsymbol{r}'_i}{\boldsymbol{r}_i}\right)\boldsymbol{r}_i(\boldsymbol{P}\widetilde{\Delta})_i^2.$$

We first consider a flow step. We note that our weights $\boldsymbol{w}$ are increasing, and if $\boldsymbol{f}''$ is increasing then $\boldsymbol{r}^{(t+1)} \geq \boldsymbol{r}^{(t)}$. Similarly if $\boldsymbol{f}''$ is decreasing, $\boldsymbol{r}^{(t+1,k)} \leq \boldsymbol{r}^{(t,k)}$. We can use the above relations to now get $\Psi(\boldsymbol{r}^{(t+1,k)}) \geq \Psi(\boldsymbol{r}^{(t,k)})$ for the first case and $\Psi(\boldsymbol{r}^{(t+1,k)}) \leq \Psi(\boldsymbol{r}^{(t,k)})$ for the second. We next consider a width reduction step. Let i be one edge that has $|\boldsymbol{P}\widetilde{\Delta}_i| \geq R\tau$. We have,

$$\boldsymbol{r}_i^{(t,k)}(\boldsymbol{P}\widetilde{\Delta})_i^2 \geq \frac{\epsilon\Phi(\boldsymbol{w}^{(t,k)})}{R^2 m}|\boldsymbol{P}\widetilde{\Delta}|_i^2 \geq \frac{\epsilon\Phi(\boldsymbol{w}^{(t,k)})}{R^2 m}R^2\tau^2 \geq \frac{\epsilon\tau^2}{(1+\epsilon)m}\Psi(\boldsymbol{r}^{(t,k)}),$$

where the last inequality follows from Lemma 2.4. Now, since we are changing our resistances by a factor of $(1+\epsilon)$, we get the following bounds for the two cases,

$$\Psi(\boldsymbol{r}^{(t,k+1)}) \geq \Psi(\boldsymbol{r}^{(t,k)}) + \left(1 - \frac{\boldsymbol{r}_i}{(1+\epsilon)\boldsymbol{r}_i}\right)\frac{\epsilon\tau^2}{(1+\epsilon)m}\Psi(\boldsymbol{r}^{(t,k)}) = \Psi(\boldsymbol{r}^{(t,k)})\left(1 + \frac{\epsilon^2\tau^2}{(1+\epsilon)^2 m}\right),$$

$$\Psi(\boldsymbol{r}^{(t,k+1)}) \leq \Psi(\boldsymbol{r}^{(t,k)}) - \frac{1}{2}\left(1 - \frac{\boldsymbol{r}_i/(1+\epsilon)}{\boldsymbol{r}_i}\right)\frac{\epsilon\tau^2}{(1+\epsilon)m}\Psi(\boldsymbol{r}^{(t,k)}) = \Psi(\boldsymbol{r}^{(t,k)})\left(1 - \frac{\epsilon^2\tau^2}{2(1+\epsilon)^2 m}\right).$$

With these two relations we conclude our proof. □

Change in Φ

Lemma 3.2. *Suppose $\boldsymbol{f}$ is M-q.s.c. Let α and τ be such that $\alpha\tau \leq M^{-1}$. After t flow steps and k width reduction steps, our potential Φ satisfies*

$$\Phi(\boldsymbol{w}^{(t,k)}) \leq \left(1 + \epsilon(1+\epsilon)^2\alpha M\right)^t\left(1 + \epsilon(1+\epsilon)\tau^{-1}\right)^k\Phi(\boldsymbol{w}_0) \quad \textit{if } \boldsymbol{f}'' \textit{ non-decreasing in } \boldsymbol{w},$$

$$\Phi(\boldsymbol{w}^{(t,k)}) \geq \left(1 - \epsilon(1+\epsilon)^2\alpha M\right)^t\left(1 - \epsilon(1+\epsilon)\tau^{-1}\right)^k\Phi(\boldsymbol{w}_0) \quad \textit{if } \boldsymbol{f}'' \textit{ non-increasing in } \boldsymbol{w}.$$

Proof. We first show the case when $\boldsymbol{f}''$ is increasing. The same calculation will work for the other case too by just considering the sign of Φ'.

We will use induction. It is easy to see the claim holds for the initial iteration, $t = k = 0$. We next assume that it holds for some $\boldsymbol{w}^{(t,k)}$. If the next step is a flow step, we update to $\boldsymbol{w}^{(t+1,k)} \leq \boldsymbol{w}^{(t,k)} + \epsilon\alpha\tau$. Since $\alpha\tau \leq M^{-1}$, we have that Φ is (M^{-1}, e^{ϵ}) hessian stable around this update. We will use $\boldsymbol{w}$ to denote $\boldsymbol{w}^{(t,k)}$ for simplicity. We thus have,

$$\begin{aligned}
\Phi(\boldsymbol{w}^{(t+1)}) =& \Phi\left(\boldsymbol{w} + \frac{\epsilon\alpha}{R}|\boldsymbol{P}\widetilde{\Delta}|\right) \\
=& \Phi(\boldsymbol{w}) + \frac{\epsilon\alpha}{R}\nabla\Phi(\boldsymbol{y})^{\top}|\boldsymbol{P}\widetilde{\Delta}| \\
& \text{(For some } \boldsymbol{y} \text{ between } \boldsymbol{w} \text{ and } \boldsymbol{w} + \alpha|\boldsymbol{P}\Delta|) \\
=& \Phi(\boldsymbol{w}) + \frac{\epsilon\alpha}{R}\sum_i \boldsymbol{f}'''(\boldsymbol{y}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \\
\leq& \Phi(\boldsymbol{w}) + \frac{\epsilon\alpha}{R}M\sum_i \boldsymbol{f}''(\boldsymbol{y}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \\
& \text{(Since } \boldsymbol{f} \text{ is } M\text{-q.s.c.)} \\
\leq& \Phi(\boldsymbol{w}) + \frac{\epsilon\alpha}{R}Me^{\epsilon}\sum_i \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \\
& \text{(Since } \boldsymbol{f} \text{ is hessian stable in this range)} \\
\leq& \Phi(\boldsymbol{w}) + \epsilon(1+\epsilon)^2\alpha M\Phi(\boldsymbol{w}) \\
& \text{(From Lemma B.3)}
\end{aligned}$$

We thus get the following bound,

$$\Phi(\boldsymbol{w}^{(t+1,k)}) \leq \Phi(\boldsymbol{w}^{(t,k)})\Big(1+\epsilon(1+\epsilon)^2\alpha M\Big).$$

Now, suppose the next step is a width reduction step.

$$\begin{aligned}
\Phi(\boldsymbol{w}^{(t,k+1)}) &= \sum_{i \notin \mathcal{I}} \Phi(\boldsymbol{w}_i) + \sum_{i \in \mathcal{I}} \Phi\Big(\boldsymbol{w}_i^{(t+1)}\Big) \\
&= \sum_{i \notin \mathcal{I}} \Phi(\boldsymbol{w}_i) + \sum_{i \in \mathcal{I}} \boldsymbol{f}''\Big(\boldsymbol{w}_i^{(t+1)}\Big) \\
&\leq \sum_{i \notin \mathcal{I}} \Phi(\boldsymbol{w}_i) + (1+\epsilon)\sum_{i \in \mathcal{I}} \boldsymbol{f}''(\boldsymbol{w}_i) \\
&\leq \Phi(\boldsymbol{w}) + \frac{\epsilon}{R\tau}\sum_{i \in \mathcal{I}} \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \\
&\leq \Phi(\boldsymbol{w}) + \frac{\epsilon}{R\tau}\sum_{i} \boldsymbol{f}''(\boldsymbol{w}_i)|\boldsymbol{P}\widetilde{\Delta}|_i \\
&\leq \Phi(\boldsymbol{w}) + \frac{\epsilon(1+\epsilon)}{\tau}\Phi(\boldsymbol{w}) \\
&\text{From Lemma B.3}
\end{aligned}$$

We thus get the following bound,

$$\Phi(\boldsymbol{w}^{(t,k+1)}) \leq \Phi(\boldsymbol{w}^{(t,k)})\Big(1+\epsilon(1+\epsilon)\tau^{-1}\Big).$$

□

B.3 Proofs from Section 4

Iterative Refinement

Lemma B.4. *Let $\boldsymbol{f}$ be a $(r, d(r))$-hessian stable function in ℓ_∞-norm, and $\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \Delta$ such that $\|\Delta\|_\infty \leq r$. We then have,*

$$\frac{1}{d(r)}\Delta^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\Delta \leq \boldsymbol{f}(\widetilde{\boldsymbol{x}}) - \boldsymbol{f}(\boldsymbol{x}) - \nabla \boldsymbol{f}(\boldsymbol{x})^\top \Delta \leq d(r)\Delta^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\Delta,$$

Proof. We have for some $\boldsymbol{z}$ along the line joining $\boldsymbol{x}$ and $\widetilde{\boldsymbol{x}}$,

$$\boldsymbol{f}(\widetilde{\boldsymbol{x}}) = \boldsymbol{f}(\boldsymbol{x}) + \nabla \boldsymbol{f}(\boldsymbol{x})^\top \Delta + \Delta^\top \nabla^2 \boldsymbol{f}(\boldsymbol{z})\Delta.$$

Since $\|\boldsymbol{z} - \boldsymbol{x}\|_\infty \leq \|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|_\infty \leq r$, from hessian stability, we have,

$$\frac{1}{d(r)}\nabla^2 \boldsymbol{f}(\boldsymbol{x}) \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{z}) \preceq d(r)\nabla^2 \boldsymbol{f}(\boldsymbol{x}).$$

Using this relation in the above, we get our lemma. □

Lemma B.5. *Let Δ be any feasible solution to the residual problem at $\boldsymbol{x}$. We then have,*

$$\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}-\Delta) \leq res(\Delta), \quad \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x} - e^{-2}\Delta) \geq e^{-2}\cdot res(\Delta),$$

Proof. Since our function is M-q.s.c., from Lemmas B.4 and B.2, for all Δ such that $\|\boldsymbol{P}\Delta\|_\infty \leq M^{-1}$,

$$e^{-1}(\boldsymbol{P}\Delta)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\boldsymbol{P}\Delta \leq \boldsymbol{f}(\boldsymbol{x}-\Delta) - \boldsymbol{f}(\boldsymbol{x}) + \nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta \leq e(\boldsymbol{P}\Delta)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\boldsymbol{P}\Delta.$$

The first bound directly follows from the left inequality. For the second bound, we first note that $e^{-2}\|\boldsymbol{P}\Delta\| \leq M^{-1}$. We can now use the right inequality.

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x} - e^{-2}\Delta) &\geq e^{-2}\nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta - e^{-3}(\boldsymbol{P}\Delta)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\boldsymbol{P}\Delta \\ &= e^{-2}\Big(\nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta - e^{-1}(\boldsymbol{P}\Delta)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x})\boldsymbol{P}\Delta\Big) \\ &= e^{-2} res(\Delta).\end{aligned}$$

□

Lemma B.6. *Assume $\boldsymbol{f}$ is M-q.s.c. Let $\boldsymbol{x}^\star$ denote the minimizer of Problem (1) and $\Delta^\star$ the optimizer of Problem (3) at $\boldsymbol{x}^{(t)}$. We then have,*

$$res(\Delta^\star) \geq \frac{1}{4MR}\Big(\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big).$$

Proof. Let $\boldsymbol{x}^{(t)}$ be such that $\boldsymbol{A}\boldsymbol{x}^{(t)} = \boldsymbol{b}$ and $\boldsymbol{x}^\star$ is the optimum of (1). Note that we have $\|\boldsymbol{P}\boldsymbol{x}^{(t)}\|_\infty \leq R$ and therefore, $\left\|\boldsymbol{P}\boldsymbol{x}^{(t)} - \boldsymbol{P}\boldsymbol{x}^\star\right\|_\infty \leq 2R$. Let $r = \frac{1}{2M}$ and $\boldsymbol{x} = \big(1 - \frac{r}{2R}\big)\boldsymbol{x}^{(t)} + \frac{r}{2R}\boldsymbol{x}^\star$. Let $\widetilde{\Delta} = \boldsymbol{x}^{(t)} - \boldsymbol{x} = \frac{r}{2R}(\boldsymbol{x}^{(t)} - \boldsymbol{x}^\star)$. We have,

$$\left\|\boldsymbol{P}\widetilde{\Delta}\right\|_\infty = \left\|\boldsymbol{P}\boldsymbol{x}^{(t)} - \boldsymbol{P}\boldsymbol{x}\right\|_\infty = \frac{r}{2R}\left\|\boldsymbol{P}\boldsymbol{x}^{(t)} - \boldsymbol{P}\boldsymbol{x}^\star\right\|_\infty \leq r,$$

and

$$\boldsymbol{A}\widetilde{\Delta} = \boldsymbol{A}(\boldsymbol{x}^{(t)} - \boldsymbol{x}) = \frac{r}{2R}(-\boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{A}\boldsymbol{x}^{(t)}) = 0.$$

We next show that $\left\|\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}\right\|_\infty \leq \frac{1}{2M}$.

$$\left\|\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}\right\|_\infty = \left\|\frac{r}{2R}\boldsymbol{P}\boldsymbol{x}^{(t)} - \frac{r}{2R}\boldsymbol{P}\boldsymbol{x}^\star - \boldsymbol{z}\right\|_\infty$$

We will do a case by case analysis. Consider some coordinate i.

1. $\boldsymbol{P}\boldsymbol{x}_i^{(t)} - \frac{1}{2M} < -R$: From the definition of $\boldsymbol{z}_i$, we note that $\boldsymbol{z}_i = R - \frac{1}{2M} + \boldsymbol{P}\boldsymbol{x}_i^{(t)}$ and $-R < \boldsymbol{P}\boldsymbol{x}_i^{(t)} \leq -R + \frac{1}{2M}$. Suppose $\boldsymbol{P}\boldsymbol{x}_i^{(t)} = -R + a$ for some $0 \leq a < \frac{1}{2M}$. We have,

$$\begin{aligned}\left|\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}\right|_i &= \left|\frac{r}{2R}(\boldsymbol{P}\boldsymbol{x}_i^{(t)} - \boldsymbol{P}\boldsymbol{x}_i^\star) - \boldsymbol{z}_i\right| \\ &= \left|\frac{r}{2R}(-R + a - \boldsymbol{P}\boldsymbol{x}_i^\star) - a + \frac{1}{2M}\right| \\ &= \left|\frac{r}{2R}(-R - \boldsymbol{P}\boldsymbol{x}_i^\star) - a\left(1 - \frac{r}{2R}\right) + \frac{1}{2M}\right| \\ &\leq \frac{1}{2M}.\end{aligned}$$

   The last inequality follows since $-2R \leq -R - \boldsymbol{P}\boldsymbol{x}_i^\star \leq 0$.

2. $\boldsymbol{P}\boldsymbol{x}_i^{(t)} + \frac{1}{2M} > R$: From the definition of $\boldsymbol{z}_i$, we note that $\boldsymbol{z}_i = -R + \frac{1}{2M} + \boldsymbol{P}\boldsymbol{x}_i^{(t)}$ and $R - \frac{1}{2M} < \boldsymbol{P}\boldsymbol{x}_i^{(t)} \leq R$. Suppose $\boldsymbol{P}\boldsymbol{x}_i^{(t)} = R - a$ for some $0 \leq a < \frac{1}{2M}$. We have,

$$\begin{aligned}\left|\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}\right|_i &= \left|\frac{r}{2R}(\boldsymbol{P}\boldsymbol{x}_i^{(t)} - \boldsymbol{P}\boldsymbol{x}_i^\star) - \boldsymbol{z}_i\right| \\ &= \left|\frac{r}{2R}(R - a - \boldsymbol{P}\boldsymbol{x}_i^\star) + a - \frac{1}{2M}\right| \\ &= \left|\frac{r}{2R}(R - \boldsymbol{P}\boldsymbol{x}_i^\star) + a\left(1 - \frac{r}{2R}\right) - \frac{1}{2M}\right| \\ &\leq \frac{1}{2M}.\end{aligned}$$

   The last inequality follows since $0 \leq R - \boldsymbol{P}\boldsymbol{x}_i^\star \leq 2R$.

3. $-R + \frac{1}{2M} \leq \boldsymbol{P}\boldsymbol{x}_i^{(t)} \leq -\frac{1}{2M}R$: In this case $\boldsymbol{z}_i = 0$.

$$\left|\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}\right|_i = \left|\frac{r}{2R}(\boldsymbol{P}\boldsymbol{x}_i^{(t)} - \boldsymbol{P}\boldsymbol{x}_i^\star)\right| \leq r = \frac{1}{2M}.$$

We thus conclude, that $\boldsymbol{x} - \boldsymbol{x}^{(t)}$ is a feasible solution for the residual problem and from convexity,

$$\frac{r}{2R}\Big(\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big) \leq \boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}).$$

Let $\Delta^\star$ denote the optimum of the residual problem at $\boldsymbol{x}^{(t)}$ (3). From Lemma B.5,

$$\frac{r}{2R}\Big(\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big) \leq \boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}) \leq res\Big(\boldsymbol{x}^{(t)} - \boldsymbol{x}\Big) \leq res(\Delta^\star).$$

□

Lemma 4.2. *[Iterative Refinement] Let $\boldsymbol{f}$ be M-q.s.c. and $\widetilde{\Delta}^{(t)}$ a κ-approximate solution to the residual problem at $\boldsymbol{x}^{(t)}$ (Problem (3)). Starting from $\boldsymbol{x}^{(0)}$ such that $\boldsymbol{A}\boldsymbol{x}^{(0)} = \boldsymbol{b}$, $\|\boldsymbol{x}^{(0)}\|_\infty \leq R$, and iterating as $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} - e^{-2}\widetilde{\Delta}^{(t)}$, after at most $O\left(\kappa M R \log\left(\frac{\boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{f}(\boldsymbol{x}^\star)}{\epsilon}\right)\right)$ iterations we get $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{x}^\star) + \epsilon$.*

Proof. From Lemma B.6,

$$res(\widetilde{\Delta}^{(t)}) \geq \frac{1}{\kappa} res(\Delta^\star) \geq \frac{1}{4\kappa M R}\Big(\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big).$$

Now, from Lemma B.5,

$$\boldsymbol{f}(\boldsymbol{x}^{(t+1)}) - \boldsymbol{f}(\boldsymbol{x}^\star) \leq \boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star) - e^{-2} res(\widetilde{\Delta}^{(t)}) \leq \left(1 - \frac{e^{-2}}{4\kappa M R}\right)\Big(\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big).$$

Inductively applying the above equation,

$$\boldsymbol{f}(\boldsymbol{x}^{(T)}) - \boldsymbol{f}(\boldsymbol{x}^\star) \leq \left(1 - \frac{e^{-2}}{4\kappa M R}\right)^T \Big(\boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{f}(\boldsymbol{x}^\star)\Big).$$

□

Binary Search

Lemma 4.3. *Let ν be such that $\boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star) \in (\nu/2, \nu]$ and $\Delta^\star$ denote the optimum of the residual problem at $\boldsymbol{x}^{(t)}$. Then, $res(\Delta^\star) \in \left(\frac{\nu}{8MR}, e^2\nu\right]$.*

Proof. The lower bound follows form B.6. For the upper bound, from B.5,

$$\nu \geq \boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x}^\star) \geq \boldsymbol{f}(\boldsymbol{x}^{(t)}) - \boldsymbol{f}(\boldsymbol{x} - e^{-2}\Delta^\star) \geq e^{-2} res(\Delta^\star).$$

□

Lemma 4.4. *Let ζ be such that $res(\Delta^\star) \in (\zeta/2, \zeta]$ and $\Delta^\star$ the optimum of the residual problem. Then, $(\boldsymbol{P}\Delta^\star)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{P}\Delta^\star \leq e \cdot \zeta$.*

Proof. Consider scaling $\Delta^\star$ by $O(1) > \lambda > 0$. We must have,

$$\left[\frac{d}{d\lambda} res(\lambda \Delta^\star)\right]_{\lambda=1} = 0.$$

This implies,

$$\nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta^\star - 2e^{-1}(\boldsymbol{P}\Delta^\star)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{P}\Delta^\star = 0,$$

or

$$e^{-1}(\boldsymbol{P}\Delta^\star)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{P}\Delta^\star = \nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta^\star - e^{-1}(\boldsymbol{P}\Delta^\star)^\top \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{P}\Delta^\star = res(\Delta^\star) \leq \zeta.$$

□

Width Reduction

Lemma 4.5. *Let ζ be such that $res(\Delta^\star) \in (\zeta/2, \zeta]$. Algorithm 3 returns $\boldsymbol{y}$ such that $\boldsymbol{A}\boldsymbol{y} = 0$, $\|\boldsymbol{P}\boldsymbol{y} - \boldsymbol{z}\|_\infty \leq \frac{1}{2M}$ and $res(\boldsymbol{y}) \geq \frac{1}{400} res(\Delta^\star)$ in $O(m^{1/3})$ calls to a linear system solver.*

Proof. This algorithm is basically an implementation of the width-reduced MWU algorithm from [CKM+11]. We will give a proof for completeness. For the purpose of this proof, we denote,

$$\Psi(\boldsymbol{r}) = \min_{\boldsymbol{A}\Delta=0, \nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\Delta = \zeta/2} \quad \sum_j \left(\boldsymbol{f}''(\boldsymbol{x}_j)(\boldsymbol{P}\Delta)_j^2 + \sum_j 4M^2 \left(\boldsymbol{w}_j + \frac{\|\boldsymbol{w}\|_1}{m} \right) \right) (\boldsymbol{P}\Delta - \boldsymbol{z})_j^2,$$

$$\Phi(\boldsymbol{w}) = \|\boldsymbol{w}\|_1.$$

Let $\widetilde{\Delta}$ be the solution returned by Ψ. We first note that, for $\Delta^\star$ the optimum of the residual problem,

$$\begin{aligned}
\Psi(\boldsymbol{r}) &\leq \sum_j \left(\boldsymbol{f}''(\boldsymbol{x}_j)(\boldsymbol{P}\Delta^\star)_j^2 + \sum_j 4M^2 \left(\boldsymbol{w}_j + \frac{\|\boldsymbol{w}\|_1}{m} \right) \right) (\boldsymbol{P}\Delta^\star - \boldsymbol{z})_j^2 \\
&\leq e \cdot \zeta + \sum_j 4M^2 \left(\boldsymbol{w}_j + \frac{\|\boldsymbol{w}\|_1}{m} \right) (\boldsymbol{P}\Delta^\star - \boldsymbol{z})_j^2, \text{ From Lemma 4.4} \\
&\leq e \cdot \zeta + \|\boldsymbol{w}\|_1 + \Phi(\boldsymbol{w}), \text{ Since } \|\boldsymbol{P}\Delta^\star - \boldsymbol{z}\|_\infty \leq \frac{1}{2M} \\
&\leq (e+2)\Phi(\boldsymbol{w}).
\end{aligned}$$

We note that,

$$\sum_j \boldsymbol{w}_j (4M)(\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z})_j \leq \sqrt{\sum_j \boldsymbol{w}_j \sum_j \boldsymbol{w}_j (4M)^2 (\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z})_j^2} \leq \sqrt{\Phi(\boldsymbol{w})\Psi(\boldsymbol{r})} \leq \sqrt{e+2}\Phi(\boldsymbol{w}). \tag{4}$$

For a flow step, from the above calculation, note that,

$$\Phi(\boldsymbol{w}^{(t+1)}) = \sum_j \boldsymbol{w}_j + \frac{\alpha}{2} \sum_j \boldsymbol{w}_j M (\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z})_j \leq \Phi(\boldsymbol{w}^{(t)}) + \frac{\sqrt{e+2}}{8} \alpha \Phi(\boldsymbol{w}^{(t)}) = \Phi(\boldsymbol{w}^{(t)})(1+\alpha).$$

For a width reduction step let $\mathcal{I}$ denote the indices which have the weights doubled,

$$\begin{aligned}
\Phi(\boldsymbol{w}^{(t+1)}) = \sum_{j \notin \mathcal{I}} \boldsymbol{w}_j^{(t)} + 2 \sum_{j \in \mathcal{I}} \boldsymbol{w}_j^{(t)} &\leq \Phi(\boldsymbol{w}^{(t)}) + \frac{2}{\tau} \sum_{j \in \mathcal{I}} \boldsymbol{w}_j^{(t)} (2M) |\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}|_j \\
&\leq \Phi(\boldsymbol{w}^{(t)}) + \frac{\sqrt{e+2}}{\tau} \Phi(\boldsymbol{w}) \leq \Phi(\boldsymbol{w}^{(t)} \left(1 + 3\tau^{-1}\right).
\end{aligned}$$

We can bound the number of width reduction steps by $O(m/\tau^2)$ similar to Lemma 3.1. We now show that our final solution has $\|\frac{1}{T}\boldsymbol{P}\boldsymbol{y} - \boldsymbol{z}\|_\infty \leq \frac{1}{2M}$. After T iterations, let j denote the index with max value in vector $\boldsymbol{w}$. For $\alpha\tau \leq 1$, $\left(1 + \frac{\alpha}{2} M |\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}|_j\right) \geq \exp\left(\frac{3}{4}\alpha M |\boldsymbol{P}\widetilde{\Delta} - \boldsymbol{z}|_j\right)$.

$$\begin{aligned}
10\zeta \geq \Phi(\boldsymbol{w}^T) \geq \boldsymbol{w}_j^{(T)} &\geq \frac{\zeta}{m} \Pi_{t=1}^T \left(1 + \frac{\alpha}{2} M |\boldsymbol{P}\widetilde{\Delta}^{(t)} - \boldsymbol{z}|_j\right) \\
&\geq \frac{\zeta}{m} \exp\left(\frac{3}{8}\alpha(2M) \sum_t |\boldsymbol{P}\widetilde{\Delta}^{(t)} - \boldsymbol{z}|_j\right) = \frac{\zeta}{m} \exp\left(\frac{3}{8}\alpha(2M)(\boldsymbol{P}\boldsymbol{y} - T\boldsymbol{z})_j\right).
\end{aligned}$$

We thus have for all coordinates j and $T \geq \alpha^{-1} O(\log m)$,

$$\frac{|\boldsymbol{P}\boldsymbol{y} - T\boldsymbol{z}|_j}{T} \leq \frac{O(M^{-1}\log m)}{\alpha T} \leq \frac{1}{2M}.$$

It remains to show that $\boldsymbol{y}/(100T)$ has the required value for the residual. First note that,

$$\nabla \boldsymbol{f}(\boldsymbol{x})^\top \frac{\boldsymbol{y}}{100T} = \frac{1}{100T} \sum_t \nabla \boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{P}\widetilde{\Delta}^{(t)} = \frac{\zeta}{2 \cdot 100}.$$

We next look at the quadratic term.

$$\frac{1}{(100)^2 T^2}\sum_j \boldsymbol{f}''(\boldsymbol{x}_j)\boldsymbol{y}_j^2 = \frac{1}{T^2(100)^2}\sum_j \boldsymbol{f}''(\boldsymbol{x}_j)\left(\sum_t |\boldsymbol{P}\widetilde{\Delta}^{(t)}|_j\right)^2$$
$$\leq \frac{1}{T^2(100)^2}\sum_j T\sum_t \boldsymbol{f}''(\boldsymbol{x}_j)|\boldsymbol{P}\widetilde{\Delta}^{(t)}|_j^2 = \frac{1}{T(100)^2}\sum_t \Psi(\boldsymbol{r}^{(t)})$$
$$\leq \frac{1}{T(100)^2}T(e+2)\Phi(\boldsymbol{w}^{(T)}) \leq \frac{10(e+2)}{(100)^2}\zeta.$$

Choose c such that we have,

$$e^{-1}\frac{1}{(100)^2}\sum_j \boldsymbol{f}''(\boldsymbol{x}_j)\boldsymbol{y}_j^2 \leq \frac{\zeta}{4\cdot 100}.$$

We thus have,

$$res\left(\frac{\boldsymbol{y}}{100T}\right) = \nabla \boldsymbol{f}(\boldsymbol{x})^\top \frac{\boldsymbol{y}}{100T} - e^{-1}\frac{1}{(100)^2T^2}\sum_j \boldsymbol{f}''(\boldsymbol{x}_j)\boldsymbol{y}_j^2 \geq \frac{\zeta}{4\cdot 100} \geq \frac{1}{400}res(\Delta^\star).$$

□

B.4 Proofs from Section 5

Sum of exponential, soft-max and ℓ_∞ regression

Theorem 5.2. *Let $\boldsymbol{x}^\star$ denote the optimum of the ℓ_∞-regression problem, $\min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \|\boldsymbol{P}\boldsymbol{x}\|_\infty$. Algorithm 1 when applied to the function $\boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}) = \sum_i \left(e^{\frac{(\boldsymbol{P}\boldsymbol{x})_i}{\nu}} + e^{\frac{-(\boldsymbol{P}\boldsymbol{x})_i}{\nu}}\right)$ for $\nu = \Omega\left(\frac{\epsilon}{\log m}\right)$, returns $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and*

$$\|\boldsymbol{P}\widetilde{\boldsymbol{x}}\|_\infty \leq (1+\epsilon)\|\boldsymbol{P}\boldsymbol{x}^\star\|_\infty,$$

in at most $\widetilde{O}(m^{1/3}\epsilon^{-5/3})$ calls to a linear system solve.

Proof. Let $\boldsymbol{Q} = \begin{bmatrix} P \\ -P \end{bmatrix}$. We note that $\boldsymbol{f}(\boldsymbol{x}) = \sum_i e^{\frac{(\boldsymbol{Q}\boldsymbol{x})_i}{\nu}}$. Let $\overline{\boldsymbol{x}}$ denote the optimum of $\boldsymbol{f}$, which is also the optimum of $smax_\nu(\boldsymbol{Q}\boldsymbol{x})$. We have the following relation,

$$\forall \boldsymbol{x}, \|\boldsymbol{P}\boldsymbol{x}\|_\infty \leq smax_\nu(\boldsymbol{Q}\boldsymbol{x}) \leq \|\boldsymbol{P}\boldsymbol{x}\|_\infty + \nu\log m.$$

Let $R = \|\boldsymbol{P}\boldsymbol{x}^\star\|_\infty$ (we can find this up to ϵ error using binary search), then the above relation implies $smax_\nu(\boldsymbol{Q}\overline{\boldsymbol{x}}) \leq R(1+\epsilon)$. From Theorem 5.1,

$$\|\boldsymbol{P}\widetilde{\boldsymbol{x}}\|_\infty \leq smax_\nu(\boldsymbol{Q}\widetilde{\boldsymbol{x}}) \leq R(1+\epsilon) = \|\boldsymbol{P}\boldsymbol{x}^\star\|_\infty(1+\epsilon).$$ □

Theorem 5.3. *For $\delta > 0$, let $\overline{\boldsymbol{x}}$ be the solution returned by Algorithm 1 (with $\epsilon = 1$) applied to $\boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}) = \sum_i e^{\frac{(\boldsymbol{P}\boldsymbol{x})_i}{\nu}}$. Now, Algorithm 2 with starting solution $\boldsymbol{x}^{(0)} = \overline{\boldsymbol{x}}$, applied to $\boldsymbol{f}$ finds $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and $\sum_i e^{\frac{(\boldsymbol{P}\widetilde{\boldsymbol{x}})_i}{\nu}} \leq (1+\delta)\sum_i e^{\frac{(\boldsymbol{P}\boldsymbol{x}^\star)_i}{\nu}}$ in at most $O\left(m^{1/3}R^2\nu^{-2}\log(\frac{m}{\delta})\right)$ calls to a linear system solver.*

Proof. From Lemma 3.3, Algorithm 1 returns $\overline{\boldsymbol{x}}$ in $O(m^{1/3})$ iterations such that $\boldsymbol{A}\overline{\boldsymbol{x}} = \boldsymbol{b}$ and $\|\boldsymbol{P}\overline{\boldsymbol{x}}\|_\infty \leq MR\|\boldsymbol{w}^{(T,K)}\|_\infty$. Since $\frac{1}{\nu^2}\sum_i e^{\frac{\boldsymbol{w}_i^{(T,K)}}{\nu}} = \Phi(\boldsymbol{w}^{(T,K)}) \leq \Phi(\boldsymbol{w}_0)e^5$, we have $\|\boldsymbol{w}^{(T,K)}\|_\infty \leq 5\nu$. This gives, $\|\boldsymbol{P}\overline{\boldsymbol{x}}\|_\infty \leq 5R$. We next bound the function value.

$$\boldsymbol{f}(\boldsymbol{P}\overline{\boldsymbol{x}}) = \sum_i e^{\frac{\boldsymbol{P}\boldsymbol{x}_i}{\nu}} \leq \sum_i e^{\frac{\boldsymbol{w}_i^{(T,K)}MR}{\nu}}.$$

If $MR \leq 1$, then $\boldsymbol{f}(\boldsymbol{P}\overline{\boldsymbol{x}}) \leq \nu^2\Phi(\boldsymbol{w}^{(T,K)}) \leq m$. Otherwise,

$$\boldsymbol{f}(\boldsymbol{P}\overline{\boldsymbol{x}}) \leq \sum_i \left(e^{\frac{\boldsymbol{w}_i^{(T,K)}}{\nu}}\right)^{MR} \leq \left(\sum_i e^{\frac{\boldsymbol{w}_i^{(T,K)}}{\nu}}\right)^{MR} \leq (\nu^2\Phi(\boldsymbol{w}^{(T,K)}))^{MR} \leq O(m^{MR}).$$

Now, we use Algorithm 2. Using the above calculated bounds in Theorem 4.6, we get our result. □

ℓ_p-Regression

Theorem 5.4. *For $\delta > 0$ and $p \geq 3$, let $\overline{\boldsymbol{x}}$ be the solution returned by Algorithm 1 (with $\epsilon = 1$) applied to $\boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}) = \|\boldsymbol{P}\boldsymbol{x}\|_p^p + \mu\|\boldsymbol{P}\boldsymbol{x}\|_2^2$. Now, Algorithm 2 with starting solution $\boldsymbol{x}^{(0)} = \overline{\boldsymbol{x}}$, applied to $\boldsymbol{f}$ finds $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and $\boldsymbol{f}(\boldsymbol{P}\widetilde{\boldsymbol{x}}) \leq \boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}^\star) + \delta$ in at most $O\left(p^2\mu^{-1/(p-2)}m^{1/3}R\log\left(\frac{pmR}{\mu\delta}\right)\right)$ calls to a linear system solver.*

Proof. From Lemma 3.3, we get $\overline{\boldsymbol{x}}$ such that $\|\overline{\boldsymbol{x}}\|_\infty \leq RM\|\boldsymbol{w}^{(T,K)}\|_\infty$. We now want to bound $\boldsymbol{f}(\overline{\boldsymbol{x}})$.

$$\boldsymbol{f}(\overline{\boldsymbol{x}}) = (RM)^p\|\boldsymbol{w}^{(T,K)}\|_p^p + \mu(RM)^2\|\boldsymbol{w}^{(T,K)}\|_2^2.$$

We next note that for $\boldsymbol{w}^{(T,K)} \geq \boldsymbol{w}_0 = 1$,

$$\Phi(\boldsymbol{w}^{(T,K)}) = p(p-1)\|\boldsymbol{w}^{(T,K)}\|_{p-2}^{p-2} + 2\mu \leq \Phi(\boldsymbol{w}_0)e^{O(1)}.$$

This implies that $\boldsymbol{w}^{(T,K)} \leq O(1)\boldsymbol{w}_0$ and $\|\boldsymbol{w}^{(T,K)}\|_\infty \leq O(1)$. Therefore,

$$\boldsymbol{f}(\overline{\boldsymbol{x}}) \leq \left((O(1)RM\right)^p m.$$

Now, using this bound on $\boldsymbol{f}(\overline{\boldsymbol{x}})$ and $\overline{\boldsymbol{x}}$ as a starting solution for Algorithm 2, we get our result by applying Theorem 4.6. □

B.4.1 Logistic Regression

Theorem 5.5. *For $\delta > 0$, let $\overline{\boldsymbol{x}}$ be the solution returned by Algorithm 1 (with $\epsilon = 1$) applied to $\boldsymbol{f}(\boldsymbol{P}\boldsymbol{x}) = \sum_i \log(1 + e^{(\boldsymbol{P}\boldsymbol{x})_i})$. Now, Algorithm 2 with starting solution $\boldsymbol{x}^{(0)} = \overline{\boldsymbol{x}}$, applied to $\boldsymbol{f}$ finds $\widetilde{\boldsymbol{x}}$ such that $\boldsymbol{A}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$ and $\sum_i \log(1 + e^{(\boldsymbol{P}\widetilde{\boldsymbol{x}})_i}) \leq \sum_i \log(1 + e^{(\boldsymbol{P}\boldsymbol{x}^\star)_i}) + \delta$ in at most $O\left(m^{1/3}R\log\left(\frac{mR}{\delta}\right)\right)$ calls to a linear system solver.*

Proof. From Lemma 3.3, we get $\overline{\boldsymbol{x}}$ such that $\|\overline{\boldsymbol{x}}\|_\infty \leq RM\|\boldsymbol{w}^{(T,K)}\|_\infty$. We now want to bound $\boldsymbol{f}(\overline{\boldsymbol{x}})$.

$$\boldsymbol{f}(\overline{\boldsymbol{x}}) = \sum_i \log(1 + e^{RM\boldsymbol{w}_i^{(T,K)}}) \leq 2RM\sum_i \boldsymbol{w}_i^{(T,K)}.$$

We next note that for $\boldsymbol{w}^{(T,K)} \geq \boldsymbol{w}_0$,

$$\Phi(\boldsymbol{w}^{(T,K)}) = \sum_i \frac{e^{\boldsymbol{w}_i^{(T,K)}}}{(1 + e^{\boldsymbol{w}_i^{(T,K)}})^2} \geq \Phi(\boldsymbol{w}_0)e^{-O(1)}.$$

This implies that $\boldsymbol{w}^{(T,K)} \leq O(1)\boldsymbol{w}_0$. Therefore,

$$\boldsymbol{f}(\overline{\boldsymbol{x}}) \leq O(Rm).$$

Now, using this bound on $\boldsymbol{f}(\overline{\boldsymbol{x}})$ and $\overline{\boldsymbol{x}}$ as a starting solution for Algorithm 2, we get our result by applying Theorem 4.6. □

C Energy Lemma

Lemma C.1. *Let $\widetilde{\Delta} = \arg\min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{c}} \boldsymbol{x}^\top\boldsymbol{P}^\top\boldsymbol{R}\boldsymbol{P}\boldsymbol{x}$. Then one has for any $\boldsymbol{r}$ and $\boldsymbol{r}'$ such that $\boldsymbol{r}' \leq \boldsymbol{r}$,*

$$\Psi(\boldsymbol{r}') \leq \Psi(\boldsymbol{r}) - \frac{1}{2}\sum_i\left(1 - \frac{\boldsymbol{r}'_i}{\boldsymbol{r}_i}\right)\boldsymbol{r}_i(\boldsymbol{P}\widetilde{\Delta})_i.$$

Proof.

$$\Psi(\boldsymbol{r}) = \min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{c}} \boldsymbol{x}^\top\boldsymbol{P}^\top\boldsymbol{R}\boldsymbol{P}\boldsymbol{x}.$$

Constructing the Lagrangian and noting that strong duality holds,

$$\begin{aligned}\Psi(\boldsymbol{r}) &= \min_{\boldsymbol{x}} \max_{\boldsymbol{y}} \quad \boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{y}^\top(\boldsymbol{c} - \boldsymbol{A}\boldsymbol{x}) \\ &= \max_{\boldsymbol{y}} \min_{\boldsymbol{x}} \quad \boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{y}^\top(\boldsymbol{c} - \boldsymbol{A}\boldsymbol{x}).\end{aligned}$$

Optimality conditions with respect to $\boldsymbol{x}$ give us,

$$2\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\boldsymbol{x}^\star = 2\boldsymbol{A}^\top \boldsymbol{y}.$$

Substituting this in Ψ gives us,

$$\Psi(\boldsymbol{r}) = \max_{\boldsymbol{y}} \quad 2\boldsymbol{y}^\top \boldsymbol{c} - \boldsymbol{y}^\top \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \boldsymbol{y}.$$

Optimality conditions with respect to $\boldsymbol{y}$ now give us,

$$2\boldsymbol{c} = 2\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \boldsymbol{y}^\star,$$

which upon re-substitution gives,

$$\Psi(\boldsymbol{r}) = \boldsymbol{c}^\top \left(\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top\right)^{-1}\boldsymbol{c}.$$

We also note that

$$\boldsymbol{x}^\star = \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \left(\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top\right)^{-1}\boldsymbol{c}. \tag{5}$$

We now want to see what happens when we change $\boldsymbol{r}$. Let $\boldsymbol{R}$ denote the diagonal matrix with entries $\boldsymbol{r}$ and let $\boldsymbol{R}' = \boldsymbol{R} - \boldsymbol{S}$, where $\boldsymbol{S}$ is the diagonal matrix with the changes in the resistances. We will use the following version of the Sherman-Morrison-Woodbury formula multiple times,

$$(\boldsymbol{X} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1}.$$

We begin by applying the above formula for $\boldsymbol{X} = \boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}$, $\boldsymbol{C} = -\boldsymbol{I}$, $\boldsymbol{U} = \boldsymbol{P}^\top \boldsymbol{S}^{1/2}$ and $\boldsymbol{V} = \boldsymbol{S}^{1/2}\boldsymbol{P}$. We thus get,

$$\begin{aligned}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1} = \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1} + \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \\ \left(\boldsymbol{I} - \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}\right)^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}.\end{aligned} \tag{6}$$

We next observe that,

$$\boldsymbol{I} - \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \preceq \boldsymbol{I},$$

which gives us,

$$\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1} \succeq \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1} + \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}. \tag{7}$$

This further implies,

$$\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \succeq \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top + \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top. \tag{8}$$

We apply the Sherman-Morrison formula again for, $\boldsymbol{X} = \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top$, $\boldsymbol{C} = \boldsymbol{I}$, $\boldsymbol{U} = \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}$ and $\boldsymbol{V} = \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top$. Let us look at the term $\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}$.

$$\begin{aligned}\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U} &= \boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top(\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \\ &\preceq \boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \\ &\preceq \boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2}.\end{aligned}$$

Using this, we get,

$$\left(\boldsymbol{A}\left(\boldsymbol{P}^{\top}\boldsymbol{R}'\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\right)^{-1} \preceq \boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1},$$

which on multiplying by $\boldsymbol{c}^{\top}$ and $\boldsymbol{c}$ gives,

$$\Psi(\boldsymbol{r}') \leq \Psi(\boldsymbol{r}) - \boldsymbol{c}^{\top}\boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{c}.$$

We note from Equation (5) that $\boldsymbol{x}^{\star} = \left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{X}^{-1}\boldsymbol{c}$. We thus have,

$$\begin{aligned}
\Psi(\boldsymbol{r}') &\leq \Psi(\boldsymbol{r}) - (\boldsymbol{x}^{\star})^{\top}\boldsymbol{P}^{\top}\boldsymbol{S}^{1/2}(\boldsymbol{I}+\boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\boldsymbol{x}^{\star} \\
&= \Psi(\boldsymbol{r}) - \sum_{e}(\boldsymbol{r}_e - \boldsymbol{r}'_e)\left(1+\frac{\boldsymbol{r}_e-\boldsymbol{r}'_e}{\boldsymbol{r}_e}\right)^{-1}(\boldsymbol{P}\boldsymbol{x}^{\star})_e \\
&= \Psi(\boldsymbol{r}) - \sum_{e}\left(\frac{\boldsymbol{r}_e-\boldsymbol{r}'_e}{2\boldsymbol{r}_e-\boldsymbol{r}'_e}\right)\boldsymbol{r}_e(\boldsymbol{P}\boldsymbol{x}^{\star})_e \\
&\leq \Psi(\boldsymbol{r}) - \frac{1}{2}\sum_{e}\left(\frac{\boldsymbol{r}_e-\boldsymbol{r}'_e}{\boldsymbol{r}_e}\right)\boldsymbol{r}_e(\boldsymbol{P}\boldsymbol{x}^{\star})_e
\end{aligned}$$

Where the last line follows from the fact $2\boldsymbol{r}_e - \boldsymbol{r}'_e \leq 2\boldsymbol{r}_e$. □

The next lemma is Lemma C.4 in [ABKS21] which is included here for completeness.

Lemma C.2. *Let* $\widetilde{\Delta} = \arg\min_{\boldsymbol{A}\boldsymbol{x}=c}\boldsymbol{x}^{\top}\boldsymbol{P}^{\top}R\boldsymbol{P}\boldsymbol{x}$. *Then one has for any* $\boldsymbol{r}'$ *and* $\boldsymbol{r}$ *such that* $\boldsymbol{r}' \geq \boldsymbol{r}$,

$$\Psi(\boldsymbol{r}') \geq \Psi(\boldsymbol{r}) + \sum_{e}\left(1-\frac{\boldsymbol{r}_e}{\boldsymbol{r}'_e}\right)\boldsymbol{r}_e(\boldsymbol{P}\widetilde{\Delta})_e^2.$$

Proof.

$$\Psi(\boldsymbol{r}) = \min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{c}} \boldsymbol{x}^{\top}\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\boldsymbol{x}.$$

Constructing the Lagrangian and noting that strong duality holds,

$$\begin{aligned}
\Psi(\boldsymbol{r}) &= \min_{\boldsymbol{x}}\max_{\boldsymbol{y}} \quad \boldsymbol{x}^{\top}\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{y}^{\top}(\boldsymbol{c}-\boldsymbol{A}\boldsymbol{x}) \\
&= \max_{\boldsymbol{y}}\min_{\boldsymbol{x}} \quad \boldsymbol{x}^{\top}\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{y}^{\top}(\boldsymbol{c}-\boldsymbol{A}\boldsymbol{x}).
\end{aligned}$$

Optimality conditions with respect to $\boldsymbol{x}$ give us,

$$2\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\boldsymbol{x}^{\star} = 2\boldsymbol{A}^{\top}\boldsymbol{y}.$$

Substituting this in Ψ gives us,

$$\Psi(\boldsymbol{r}) = \max_{\boldsymbol{y}} \quad 2\boldsymbol{y}^{\top}\boldsymbol{c} - \boldsymbol{y}^{\top}\boldsymbol{A}\left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{y}.$$

Optimality conditions with respect to $\boldsymbol{y}$ now give us,

$$2\boldsymbol{c} = 2\boldsymbol{A}\left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{y}^{\star},$$

which upon re-substitution gives,

$$\Psi(\boldsymbol{r}) = \boldsymbol{c}^{\top}\left(\boldsymbol{A}\left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\right)^{-1}\boldsymbol{c}.$$

We also note that

$$\boldsymbol{x}^{\star} = \left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\left(\boldsymbol{A}\left(\boldsymbol{P}^{\top}\boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^{\top}\right)^{-1}\boldsymbol{c}. \tag{9}$$

We now want to see what happens when we change $\boldsymbol{r}$. Let $\boldsymbol{R}$ denote the diagonal matrix with entries $\boldsymbol{r}$ and let $\boldsymbol{R}' = \boldsymbol{R} + \boldsymbol{S}$, where $\boldsymbol{S}$ is the diagonal matrix with the changes in the resistances. We will use the following version of the Sherman-Morrison-Woodbury formula multiple times,

$$(\boldsymbol{X} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1}.$$

We begin by applying the above formula for $\boldsymbol{X} = \boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}$, $\boldsymbol{C} = \boldsymbol{I}$, $\boldsymbol{U} = \boldsymbol{P}^\top \boldsymbol{S}^{1/2}$ and $\boldsymbol{V} = \boldsymbol{S}^{1/2}\boldsymbol{P}$. We thus get,

$$\begin{aligned}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1} = \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1} - \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \\ \left(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}\right)^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}. \quad (10)\end{aligned}$$

We next claim that

$$\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2} \preceq \boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2},$$

which gives us,

$$\begin{aligned}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1} \preceq \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1} - \\ \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}. \quad (11)\end{aligned}$$

This further implies,

$$\begin{aligned}\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \preceq \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top - \\ \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top. \quad (12)\end{aligned}$$

We apply the Sherman-Morrison formula again for, $\boldsymbol{X} = \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top$, $\boldsymbol{C} = -(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}$, $\boldsymbol{U} = \boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^\top \boldsymbol{S}^{1/2}$ and $\boldsymbol{V} = \boldsymbol{S}^{1/2}\boldsymbol{P}\left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top$. Let us look at the term $\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}$.

$$-\left(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}\right)^{-1} = \left(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2} - \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}\right)^{-1} \succeq (\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}.$$

Using this, we get,

$$\left(\boldsymbol{A}\left(\boldsymbol{P}^\top \boldsymbol{R}'\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top\right)^{-1} \succeq \boldsymbol{X}^{-1} + \boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1},$$

which on multiplying by $\boldsymbol{c}^\top$ and $\boldsymbol{c}$ gives,

$$\Psi(\boldsymbol{r}') \geq \Psi(\boldsymbol{r}) + \boldsymbol{c}^\top \boldsymbol{X}^{-1}\boldsymbol{U}(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{c}.$$

We note from Equation (9) that $\boldsymbol{x}^\star = \left(\boldsymbol{P}^\top \boldsymbol{R}\boldsymbol{P}\right)^{-1}\boldsymbol{A}^\top \boldsymbol{X}^{-1}\boldsymbol{c}$. We thus have,

$$\begin{aligned}\Psi(\boldsymbol{r}') &\geq \Psi(\boldsymbol{r}) + (\boldsymbol{x}^\star)^\top \boldsymbol{P}^\top \boldsymbol{S}^{1/2}(\boldsymbol{I} + \boldsymbol{S}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{S}^{1/2})^{-1}\boldsymbol{S}^{1/2}\boldsymbol{P}\boldsymbol{x}^\star \\ &= \Psi(\boldsymbol{r}) + \sum_e \left(\frac{\boldsymbol{r}'_e - \boldsymbol{r}_e}{\boldsymbol{r}'_e}\right)\boldsymbol{r}_e(\boldsymbol{P}\boldsymbol{x}^\star)_e.\end{aligned}$$

□