First-order Potential Reduction Method

Project Notes

September 21, 2022

1 Dimension-reduced Method for Potential Reduction

In this section, we discuss the application of dimension-reduced method to potential reduction. For brevity, we for now only consider the primal potential reduction and focus on the simplex-constrained QP.

$$\begin{aligned} \min_{\mathbf{x}} \quad & \tfrac{1}{2}\left\|\mathbf{A}\mathbf{x}\right\|^2 \quad =: f\left(\mathbf{x}\right) \\ \text{subject to} \quad & \mathbf{e}^\top \mathbf{x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

and we adopt the potential function

$$\varphi\left(\mathbf{x}\right) := \rho \log\left(f\left(\mathbf{x}\right)\right) - \sum_{i=1}^{n} \log x_i,$$

whose gradient is given by

$$\nabla\varphi\left(\mathbf{x}\right) = \frac{\rho \nabla f\left(\mathbf{x}\right)}{f\left(\mathbf{x}\right)} - \mathbf{X}^{-1}\mathbf{e}.$$

At each iteration, we evaluate the gradient $\nabla\varphi\left(\mathbf{x}^k\right)$, let $\Delta := \mathbf{x}^{k+1} - \mathbf{x}^k$ and solve following subproblem

$$\begin{aligned} \min_{\Delta} \quad & \left\langle \nabla\varphi\left(\mathbf{x}^k\right), \Delta \right\rangle \\ \text{subject to} \quad & \mathbf{e}^\top \Delta = 0 \\ & \left\| \left(\mathbf{X}^k\right)^{-1} \Delta \right\| \leq \beta. \end{aligned}$$

Starting from the basic potential reduction, we extend it by incorporating momentum term for faster convergence.

1.1 Two directions

In this section, we consider two direction extension of the potential reduction framework. In a word, by keeping track of one recent history iterate, we update

$$\begin{aligned} \mathbf{d}^k \quad & \leftarrow \quad \alpha^g \mathbf{P}_\Delta\left[\nabla\varphi\left(\mathbf{x}^k\right)\right] + \alpha^m\left(\mathbf{x}^k - \mathbf{x}^{k-1}\right) \\ \mathbf{x}^k \quad & \leftarrow \quad \mathbf{x}^k + \mathbf{d}^k \end{aligned}$$

where $\mathbf{P}_\Delta[\cdot]$ is the orthogonal projection onto $\mathbf{e}^\top\mathbf{x}=0$. Note that we compute α^g, α^d through the following model

$$\begin{array}{rl}\min\limits_{\mathbf{d},\alpha^g,\alpha^m} & \frac{1}{2}\mathbf{d}^\top\mathbf{H}\mathbf{d}+\mathbf{h}^\top\mathbf{d} \\ \text{subject to} & \left\|\mathbf{X}^{-1}\mathbf{d}\right\|\le\Delta \\ & \mathbf{d}=\alpha^g\mathbf{g}^k+\alpha^m\mathbf{m}^k\end{array}$$

where $\mathbf{g}^k := \mathbf{P}_\Delta\left[\nabla\varphi\left(\mathbf{x}^k\right)\right]$, $\mathbf{m}^k := \mathbf{x}^k-\mathbf{x}^{k-1}$. Alternatively, we define $\mathbf{G} := \begin{pmatrix} | & | \\ \mathbf{g}^k & \mathbf{m}^k \\ | & | \end{pmatrix}$, $\alpha = \begin{pmatrix}\alpha^g \\ \alpha^m\end{pmatrix}$ and $\mathbf{d}=\mathbf{G}\alpha$, giving

$$\begin{array}{rl}\min\limits_{\alpha} & \frac{1}{2}\alpha^\top\mathbf{G}^\top\mathbf{H}\mathbf{G}^\top\alpha+\mathbf{h}^\top\mathbf{G}\alpha \\ \text{subject to} & \left\|\mathbf{X}^{-1}\mathbf{G}\alpha\right\|\le\Delta,\end{array}$$

or

$$\begin{array}{rl}\min\limits_{\alpha} & \frac{1}{2}\alpha^\top\widetilde{\mathbf{H}}\alpha+\widetilde{\mathbf{h}}\alpha \quad =: m(\alpha) \\ \text{subject to} & \|\mathbf{M}\alpha\|\le\Delta\end{array}$$

for

$$\begin{aligned}\widetilde{\mathbf{H}} &:= \begin{pmatrix}\left\langle\mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{g}^k\right\rangle & \left\langle\mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{m}^k\right\rangle \\ \left\langle\mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{g}^k\right\rangle & \left\langle\mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{m}^k\right\rangle\end{pmatrix} \\ \widetilde{\mathbf{h}} &:= \begin{pmatrix}\left\|\mathbf{g}^k\right\|^2 \\ \left\langle\mathbf{g}^k,\mathbf{m}^k\right\rangle\end{pmatrix} \\ \mathbf{M} &:= \begin{pmatrix}\left\|\left(\mathbf{X}^k\right)^{-1}\mathbf{g}^k\right\|^2 & \left\langle\mathbf{g}^k,\left(\mathbf{X}^k\right)^{-2}\mathbf{m}^k\right\rangle \\ \left\langle\mathbf{m}^k,\left(\mathbf{X}^k\right)^{-2}\mathbf{g}^k\right\rangle & \left\|\left(\mathbf{X}^k\right)^{-1}\mathbf{m}^k\right\|^2\end{pmatrix}.\end{aligned}$$

Note that $\nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right) = -\frac{\rho\nabla f\left(\mathbf{x}^k\right)\nabla f\left(\mathbf{x}^k\right)^\top}{f\left(\mathbf{x}^k\right)^2}+\rho\frac{\mathbf{A}^\top\mathbf{A}}{f\left(\mathbf{x}^k\right)}+\left(\mathbf{X}^k\right)^{-2}$ and we evaluate the above relations via

$$\begin{aligned}\left\langle\mathbf{a},\nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{a}\right\rangle &= \left\langle\mathbf{a},-\frac{\rho\nabla f\left(\mathbf{x}^k\right)\nabla f\left(\mathbf{x}^k\right)^\top\mathbf{a}}{f\left(\mathbf{x}^k\right)^2}\right\rangle+\frac{\|\mathbf{A}\mathbf{a}\|^2}{f\left(\mathbf{x}^k\right)}+\left\|\left(\mathbf{X}^k\right)^{-1}\mathbf{a}\right\|^2 \\ &= -\rho\left(\frac{\nabla f\left(\mathbf{x}^k\right)^\top\mathbf{a}}{f\left(\mathbf{x}^k\right)}\right)^2+\frac{\|\mathbf{A}\mathbf{a}\|^2}{f\left(\mathbf{x}^k\right)}+\left\|\left(\mathbf{X}^k\right)^{-1}\mathbf{a}\right\|^2 \\ \left\langle\mathbf{a},\nabla^2_{\mathbf{x},\mathbf{x}}\varphi\left(\mathbf{x}^k\right)\mathbf{b}\right\rangle &= \left\langle\mathbf{a},-\frac{\rho\nabla f\left(\mathbf{x}^k\right)\nabla f\left(\mathbf{x}^k\right)^\top\mathbf{b}}{f\left(\mathbf{x}^k\right)^2}\right\rangle+\frac{\langle\mathbf{A}\mathbf{a},\mathbf{A}\mathbf{b}\rangle}{f\left(\mathbf{x}^k\right)}+\left\langle\mathbf{a},\left(\mathbf{X}^k\right)^{-2}\mathbf{b}\right\rangle \\ &= -\rho\left(\frac{\nabla f\left(\mathbf{x}^k\right)^\top\mathbf{a}}{f\left(\mathbf{x}^k\right)}\right)\left(\frac{\nabla f\left(\mathbf{x}^k\right)^\top\mathbf{b}}{f\left(\mathbf{x}^k\right)}\right)+\frac{\langle\mathbf{A}\mathbf{a},\mathbf{A}\mathbf{b}\rangle}{f\left(\mathbf{x}^k\right)}+\left\langle\mathbf{a},\left(\mathbf{X}^k\right)^{-2}\mathbf{b}\right\rangle.\end{aligned}$$

To ensure feasibility, we always choose $\Delta\le 1$ and adjust it based on the trust-region rule.

2 Potential Reduction for LP

In this section, we discuss the potential reduction method on LP HSD model.

$$\begin{array}{rc} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

$$\begin{array}{rc} \max\limits_{\mathbf{y}\in\mathbb{R}^m} & \mathbf{b}^\top \mathbf{y} \\ \text{subject to} & \mathbf{A}^\top \mathbf{y} + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \geq \mathbf{0} \end{array}$$

and

$$\begin{array}{rcl} \mathbf{A}\mathbf{x} - \mathbf{b}\tau & = & \mathbf{0} \\ -\mathbf{A}^\top \mathbf{y} - \mathbf{s} + \mathbf{c}\tau & = & \mathbf{0} \\ \mathbf{b}^\top \mathbf{y} - \mathbf{c}^\top \mathbf{x} - \kappa & = & 0 \\ \mathbf{e}_n^\top \mathbf{x} + \mathbf{e}_n^\top \mathbf{s} + \kappa + \tau & = & 1 \end{array}$$

2.1 Potential Reduction for HSD

In this section we consider the original HSD formulation

$$\begin{array}{rcl} \mathbf{A}\mathbf{x} - \mathbf{b}\tau & = & \mathbf{0} \\ -\mathbf{A}^\top \mathbf{y} - \mathbf{s} + \mathbf{c}\tau & = & \mathbf{0} \\ \mathbf{b}^\top \mathbf{y} - \mathbf{c}^\top \mathbf{x} - \kappa & = & 0 \\ \mathbf{e}_n^\top \mathbf{x} + \mathbf{e}_n^\top \mathbf{s} + \kappa + \tau & = & 1 \end{array}$$

and we have

$$\begin{array}{rcl} \begin{pmatrix} \mathbf{0}_{m\times m} & \mathbf{A} & \mathbf{0}_{m\times n} & \mathbf{0}_{m\times 1} & -\mathbf{b} \\ -\mathbf{A}^\top & \mathbf{0}_{n\times n} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} & \mathbf{c} \\ \mathbf{b}^\top & -\mathbf{c}^\top & \mathbf{0}_{1\times n} & -1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{s} \\ \kappa \\ \tau \end{pmatrix} & = & \mathbf{0} \\ \mathbf{e}_n^\top \mathbf{x} + \mathbf{e}_n^\top \mathbf{s} + \kappa + \tau & = & 1. \end{array}$$

In this method, the dual variable $\mathbf{y}$ is free and needs special treatment. First we consider the potential function

$$\begin{array}{rcl} f(\mathbf{x}, \mathbf{y}, \mathbf{s}, \kappa, \tau) & := & \frac{1}{2} \left\| \widetilde{\mathbf{A}} \mathbf{u} \right\|^2 \\ \varphi(\mathbf{x}, \mathbf{y}, \mathbf{s}, \kappa, \tau) & := & \rho \log(f(\mathbf{u})) - B(\mathbf{x}) - B(\mathbf{s}) - \log \kappa - \log \tau \end{array}$$

and

$$
\begin{aligned}
& \nabla f\left(\mathbf{x}, \mathbf{y}, \mathbf{s}, \kappa, \tau\right) \\
= \; & \widetilde{\mathbf{A}}^{\top} \widetilde{\mathbf{A}} \mathbf{u} \\
= \; & \begin{pmatrix} \mathbf{0}_{m\times m} & \mathbf{A} & \mathbf{0}_{m\times n} & \mathbf{0}_{m\times 1} & -\mathbf{b} \\ -\mathbf{A}^{\top} & \mathbf{0}_{n\times n} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} & \mathbf{c} \\ \mathbf{b}^{\top} & -\mathbf{c}^{\top} & \mathbf{0}_{1\times n} & -1 & 0 \end{pmatrix}^{\top} \begin{pmatrix} \mathbf{A}\mathbf{x} - \mathbf{b}\tau & =: \mathbf{r}_1 \\ -\mathbf{A}^{\top}\mathbf{y} - \mathbf{s} + \mathbf{c}\tau & =: \mathbf{r}_2 \\ \mathbf{b}^{\top}\mathbf{y} - \mathbf{c}^{\top}\mathbf{x} - \kappa & =: r_3 \end{pmatrix} \\
= \; & \begin{pmatrix} \mathbf{0}_{m\times m} & -\mathbf{A} & \mathbf{b} \\ \mathbf{A}^{\top} & \mathbf{0}_{n\times n} & -\mathbf{c} \\ \mathbf{0}_{n\times m} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times m} & \mathbf{0}_{1\times n} & -1 \\ -\mathbf{b}^{\top} & \mathbf{c}^{\top} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} -\mathbf{A}\mathbf{r}_2 + \mathbf{b} r_3 \\ \mathbf{A}^{\top}\mathbf{r}_1 - \mathbf{c} r_3 \\ -\mathbf{r}_2 \\ -r_3 \\ -\mathbf{b}^{\top}\mathbf{r}_1 + \mathbf{c}^{\top}\mathbf{r}_2 \end{pmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\nabla \varphi\left(\mathbf{u}\right) & = \frac{\rho \nabla f\left(\mathbf{u}\right)}{f\left(\mathbf{u}\right)} - \begin{pmatrix} \mathbf{0}_m \\ \mathbf{X}^{-1}\mathbf{e} \\ \mathbf{S}^{-1}\mathbf{e} \\ \kappa^{-1} \\ \tau^{-1} \end{pmatrix} \\
\nabla^2_{\mathbf{u},\mathbf{u}} \varphi\left(\mathbf{u}\right) & = -\frac{\rho \nabla f\left(\mathbf{u}\right) \nabla f\left(\mathbf{u}\right)^{\top}}{f\left(\mathbf{u}\right)^2} + \rho \frac{\widetilde{\mathbf{A}}^{\top}\widetilde{\mathbf{A}}}{f\left(\mathbf{u}\right)} + \operatorname{diag} \begin{pmatrix} \mathbf{0}_m \\ \mathbf{X}^{-2}\mathbf{e} \\ \mathbf{S}^{-2}\mathbf{e} \\ \kappa^{-2} \\ \tau^{-2} \end{pmatrix}.
\end{aligned}
$$

2.2 Acceleration by negative curvature

In this section, we discuss how to find the negative curvature of the Hessian to help accelerate algorithm convergence. More specifically, we consider the following problem

$$
\lambda_{\min} \left\{ \nabla^2 \varphi\left(\mathbf{u} = (\mathbf{x}, \mathbf{y})\right) = \frac{2\rho \mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^2} - \frac{4\rho \mathbf{A}^{\top}\mathbf{A}\mathbf{u}\mathbf{u}^{\top}\mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\}.
$$

And we wish to solve the eigen-problem

$$
\begin{aligned}
\min_{\|\mathbf{v}\|=1} \quad & \mathbf{v}^{\top} \left\{ \tfrac{2\rho \mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^2} - \tfrac{4\rho \mathbf{A}^{\top}\mathbf{A}\mathbf{u}\mathbf{u}^{\top}\mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\} \mathbf{v} \\
\text{subject to} \quad & \mathbf{e}^{\top}\mathbf{v}_{\mathbf{x}} = 0.
\end{aligned}
$$

In general there are two ways to compute a valid direction. The first method approaches the problem directly and uses Lanczos iteration to find the negative eigen-value of $\nabla^2\varphi$. As for the second approach, we apply the scaling matrix $\mathbf{S} := \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix}$ and solve

$$
\begin{aligned}
\min_{\|\mathbf{S}\mathbf{v}\|=1} \quad & \mathbf{v}^{\top} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \left\{ \tfrac{2\rho \mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^2} - \tfrac{4\rho \mathbf{A}^{\top}\mathbf{A}\mathbf{u}\mathbf{u}^{\top}\mathbf{A}^{\top}\mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \\
\text{subject to} \quad & \mathbf{x}^{\top}\mathbf{v}_{\mathbf{x}} = 0.
\end{aligned}
$$

To improve the conditioning of the Hessian, we replace $\|\mathbf{Sv}\| = 1$ by $\|\mathbf{v}\| = 1$ and arrive at

$$\begin{array}{ll} \min\limits_{\|\mathbf{v}\|=1} & \mathbf{v}^\top \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \left\{ \frac{2\rho\mathbf{A}^\top\mathbf{A}}{\|\mathbf{Au}\|^2} - \frac{4\rho\mathbf{A}^\top\mathbf{Auu}^\top\mathbf{A}^\top\mathbf{A}}{\|\mathbf{Au}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \\ \text{subject to} & \mathbf{x}^\top \mathbf{v}_\mathbf{x} = 0. \end{array}$$

Another trick we apply is to ignore the variables which are predicted to be nonbasic in the optimal solution so that the Hessian computation can be greatly simplified.

2.3 Direct computation

When evaluating the Hessian, it is possible that the matrix is ill-conditioned. Hence we need to consider the following relation

$$\begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n \end{pmatrix} \left[\frac{2\rho\mathbf{A}^\top\mathbf{A}}{\|\mathbf{Au}\|^2} - \frac{4\rho\mathbf{A}^\top\mathbf{Auu}^\top\mathbf{A}^\top\mathbf{A}}{\|\mathbf{Au}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right] \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n \end{pmatrix}$$

$$\left(\mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n\right) \mathbf{X}^{-2} \left(\mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n\right) = \mathbf{X}^{-2}\mathbf{v} - \frac{\mathbf{X}^{-1}\mathbf{e}_n^\top}{n}\mathbf{v} - \frac{\mathbf{e}_n\mathbf{e}_n^\top}{n}\mathbf{X}^{-1}\mathbf{v} + \frac{\mathbf{e}_n\mathbf{e}_n^\top}{n^2}\mathbf{e}_n^\top\mathbf{X}^{-2}\mathbf{e}_n$$

$$\begin{aligned} \mathbf{v} &\leftarrow \begin{pmatrix} \mathbf{v}_\mathbf{y} \\ \mathbf{v}_\mathbf{x} - \left(\mathbf{e}_n^\top\mathbf{v}\right)\mathbf{v}/n \end{pmatrix} \\ \mathbf{u}_1 &\leftarrow \begin{pmatrix} \mathbf{0} \\ \mathbf{v}_\mathbf{x} - \mathbf{e}_n\mathbf{e}_n^\top/n \end{pmatrix} \\ \mathbf{u}_2 &\leftarrow \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n \end{pmatrix} \mathbf{A}^\top\mathbf{Av} \\ \mathbf{u}_3 &\leftarrow (\mathbf{g}^\top\mathbf{v}) \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{e}_n\mathbf{e}_n^\top/n \end{pmatrix} \mathbf{g} \\ \frac{1}{4}\|\mathbf{Au}\|^4 \mathbf{Mv} &\leftarrow f^2\mathbf{u}_1 + 2\rho f\mathbf{u}_2 - 4\rho\mathbf{u}_3 \end{aligned}$$

2.4 Scaled Hessian

$$\mathbf{M} := \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{xx}^\top/\|\mathbf{x}\|^2 \end{pmatrix} \left[\frac{2\rho\mathbf{SA}^\top\mathbf{AS}}{\|\mathbf{Au}\|^2} - \frac{4\rho\mathbf{SA}^\top\mathbf{Auu}^\top\mathbf{A}^\top\mathbf{AS}}{\|\mathbf{Au}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{I}_n \end{pmatrix} \right] \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{xx}^\top/\|\mathbf{x}\|^2 \end{pmatrix}$$

In the computation of scaled Hessian, we implement matrix-vector product $\mathbf{Mv}$ as follows

$$
\begin{aligned}
\mathbf{x}' &\leftarrow \frac{\mathbf{x}}{\|\mathbf{x}\|} \\
\mathbf{v} &\leftarrow \begin{pmatrix} \mathbf{v_y} \\ \mathbf{v_x} - \left(\mathbf{x}'^\top \mathbf{v_x}\right)\mathbf{x}' \end{pmatrix} \\
\mathbf{u}_1 &\leftarrow \begin{pmatrix} \mathbf{0} \\ \mathbf{v_x} - \left(\mathbf{x}'^\top \mathbf{v_x}\right)\mathbf{x}' \end{pmatrix} \\
\mathbf{u}_2 &\leftarrow \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{x}'\mathbf{x}'^\top \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{A}^\top \mathbf{A} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \\
\mathbf{u}_3 &\leftarrow \mathbf{g}^\top \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{x}'\mathbf{x}'^\top \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{g} \\
\|\mathbf{Au}\|^4 \mathbf{Mv} &\leftarrow f^2 \mathbf{u}_1 + \rho f \mathbf{u}_2 - \rho \mathbf{u}_3
\end{aligned}
$$

2.5 Further simplification

There are some basic operations to implement

Residual setup

$$
\begin{aligned}
\mathbf{r}_1 &= \mathbf{Ax} - \mathbf{b}\tau \\
\mathbf{r}_2 &= -\mathbf{A}^\top \mathbf{y} - \mathbf{s} + \mathbf{c}\tau \\
r_3 &= \mathbf{b}^\top \mathbf{y} - \mathbf{c}^\top \mathbf{x} - \kappa.
\end{aligned}
$$

Objective value

$$
f = \frac{1}{2}\left[\|\mathbf{r}_1\|^2 + \|\mathbf{r}_2\|^2 + r_3^2\right]
$$

Gradient setup

$$
\nabla f = \begin{pmatrix} -\mathbf{A}\mathbf{r}_2 + \mathbf{b}r_3 \\ \mathbf{A}^\top \mathbf{r}_1 - \mathbf{c}r_3 \\ -\mathbf{r}_2 \\ -r_3 \\ -\mathbf{b}^\top \mathbf{r}_1 + \mathbf{c}^\top \mathbf{r}_2 \end{pmatrix}
$$

$$
\nabla \varphi = \frac{\rho \nabla f}{f} - \begin{pmatrix} \mathbf{X}^{-1}\mathbf{e} \\ \mathbf{0}_m \\ \mathbf{S}^{-1}\mathbf{e} \\ \kappa^{-1} \\ \tau^{-1} \end{pmatrix}
$$

Hessian (no actual setup)

$$\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}} = \begin{pmatrix} \mathbf{0}_{m\times m} & -\mathbf{A} & \mathbf{b} \\ \mathbf{A}^\top & \mathbf{0}_{n\times n} & -\mathbf{c} \\ \mathbf{0}_{n\times m} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times m} & \mathbf{0}_{1\times n} & -1 \\ -\mathbf{b}^\top & \mathbf{c}^\top & 0 \end{pmatrix} \begin{pmatrix} \mathbf{0}_{m\times m} & \mathbf{A} & \mathbf{0}_{m\times n} & \mathbf{0}_{m\times 1} & -\mathbf{b} \\ -\mathbf{A}^\top & \mathbf{0}_{n\times n} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} & \mathbf{c} \\ \mathbf{b}^\top & -\mathbf{c}^\top & \mathbf{0}_{1\times n} & -1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{A}\mathbf{A}^\top + \mathbf{b}\mathbf{b}^\top & -\mathbf{b}\mathbf{c}^\top & \mathbf{A} & -\mathbf{b} & -\mathbf{A}\mathbf{c} \\ -\mathbf{c}\mathbf{b}^\top & \mathbf{A}^\top\mathbf{A} & \mathbf{0}_{n\times n} & \mathbf{c} & -\mathbf{A}^\top\mathbf{b} \\ \mathbf{A}^\top & \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} & -\mathbf{c} \\ -\mathbf{b}^\top & \mathbf{c}^\top & \mathbf{0}_{1\times n} & \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times 1} \\ -\mathbf{c}^\top\mathbf{A}^\top & -\mathbf{b}^\top\mathbf{A} & -\mathbf{c}^\top & \mathbf{0}_{1\times 1} & \|\mathbf{b}\|^2 + \|\mathbf{c}\|^2 \end{pmatrix}$$

$$\nabla^2\varphi = -\frac{\rho\nabla f\nabla f^\top}{f^2} + \frac{\rho\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}}{f} + \begin{pmatrix} \mathbf{X}^{-2} & & & & \\ & \mathbf{0}_m & & & \\ & & \mathbf{S}^{-2} & & \\ & & & \kappa^{-2} & \\ & & & & \tau^{-2} \end{pmatrix}.$$

Hessian-vector (with projection)

$$\mathbf{u} = \mathbf{x} - \frac{\mathbf{e}^\top\mathbf{x}}{n}\cdot\mathbf{e}$$

$$\nabla^2\varphi\mathbf{u} = -\frac{\rho\left(\nabla f^\top\mathbf{u}\right)}{f^2}\nabla f + \frac{\rho}{f}\widetilde{\mathbf{A}}^\top\left(\widetilde{\mathbf{A}}\mathbf{u}\right) + \begin{pmatrix} \mathbf{X}^{-2} & & & & \\ & \mathbf{0}_{m\times m} & & & \\ & & \mathbf{S}^{-2} & & \\ & & & \kappa^{-2} & \\ & & & & \tau^{-2} \end{pmatrix}\mathbf{u}.$$

Minimal eigenvalue

To evaluate the minimum eigen-value of $\mathbf{P}_\Delta\nabla^2\varphi\mathbf{P}_\Delta$ and the corresponding eigen-vector

$$\mathbf{X}\nabla^2\varphi\mathbf{X} = -\frac{4\rho\mathbf{X}\mathbf{A}^\top\mathbf{A}\mathbf{x}\mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{X}}{\|\mathbf{A}\mathbf{x}\|^4} + \frac{2\rho\mathbf{X}\mathbf{A}^\top\mathbf{A}\mathbf{X}}{\|\mathbf{A}\mathbf{x}\|^2} + \mathbf{I}$$

Note that

$$\begin{aligned} & \mathbf{X}\nabla^2\varphi(\mathbf{u})\mathbf{X} \\ = \; & -\frac{\rho\mathbf{X}\nabla f(\mathbf{u})\nabla f(\mathbf{u})^\top\mathbf{X}}{f(\mathbf{u})^2} + \frac{\rho\mathbf{X}\mathbf{A}^\top\mathbf{A}\mathbf{X}}{f(\mathbf{u})} + \mathbf{D}. \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{v}} \quad & \left\langle \mathbf{X}\mathbf{v}, \nabla^2\varphi(\mathbf{u})\mathbf{X}\mathbf{v}\right\rangle \\ \text{subject to} \quad & \mathbf{e}^\top\mathbf{X}\mathbf{v} = 0 \\ & \|\mathbf{X}\mathbf{v}\| = 1 \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{v}} \quad & \left\langle \mathbf{v}, \left(\mathbf{I} - \tfrac{\mathbf{x}\mathbf{x}^\top}{\|\mathbf{x}\|^2}\right)\left(\mathbf{X}\nabla^2\varphi(\mathbf{u})\mathbf{X}\right)\left(\mathbf{I} - \tfrac{\mathbf{x}\mathbf{x}^\top}{\|\mathbf{x}\|^2}\right)\mathbf{v}\right\rangle \\ \text{subject to} \quad & \|\mathbf{X}\mathbf{v}\| = 1 \end{aligned}$$

$$
\begin{aligned}
& \left(\mathbf{I}-\frac{\mathbf{x}\mathbf{x}^{\top}}{\|\mathbf{x}\|^{2}}\right)\mathbf{H}\left(\mathbf{I}-\frac{\mathbf{x}\mathbf{x}^{\top}}{\|\mathbf{x}\|^{2}}\right) \\
= & \mathbf{H}-\frac{\mathbf{x}\mathbf{x}^{\top}}{\|\mathbf{x}\|^{2}}\mathbf{H}-\mathbf{H}\frac{\mathbf{x}\mathbf{x}^{\top}}{\|\mathbf{x}\|^{2}}+\mathbf{x}^{\top}\mathbf{H}\mathbf{x}\frac{\mathbf{x}\mathbf{x}^{\top}}{\|\mathbf{x}\|^{4}}
\end{aligned}
$$

3 Numerical Experiments

Problem	PInfeas	DInfeas.	Compl.	Problem	PInfeas	DInfeas.	Compl.
DLITTLE	1.347e-10	2.308e-10	2.960e-09	KB2	5.455e-11	6.417e-10	7.562e-11
AFIRO	7.641e-11	7.375e-11	3.130e-10	LOTFI	2.164e-09	4.155e-09	8.663e-08
AGG2	3.374e-08	4.859e-08	6.286e-07	MODSZK1	1.527e-06	5.415e-05	2.597e-04
AGG3	2.248e-05	1.151e-06	1.518e-05	RECIPELP	5.868e-08	6.300e-08	1.285e-07
BANDM	2.444e-09	4.886e-09	3.769e-08	SC105	7.315e-11	5.970e-11	2.435e-10
BEACONFD	5.765e-12	9.853e-12	1.022e-10	SC205	6.392e-11	5.710e-11	2.650e-10
BLEND	2.018e-10	3.729e-10	1.179e-09	SC50A	1.078e-05	6.098e-06	4.279e-05
BOEING2	1.144e-07	1.110e-08	2.307e-07	SC50B	4.647e-11	3.269e-11	1.747e-10
BORE3D	2.389e-08	5.013e-08	1.165e-07	SCAGR25	1.048e-07	5.298e-08	1.289e-06
BRANDY	2.702e-05	7.818e-06	1.849e-05	SCAGR7	1.087e-07	1.173e-08	2.601e-07
CAPRI	7.575e-05	4.488e-05	4.880e-05	SCFXM1	4.323e-06	5.244e-06	8.681e-06
E226	2.656e-06	4.742e-06	2.512e-05	SCORPION	1.674e-09	1.892e-09	1.737e-08
FINNIS	8.577e-07	8.367e-07	1.001e-05	SCTAP1	5.567e-07	8.430e-07	5.081e-06
FORPLAN	5.874e-07	2.084e-07	4.979e-06	SEBA	2.919e-11	5.729e-11	1.448e-10
GFRD-PNC	4.558e-05	1.052e-05	4.363e-05	SHARE1B	3.367e-07	1.339e-06	3.578e-06
GROW7	1.276e-04	4.906e-06	1.024e-04	SHARE2B	2.142e-04	2.014e-05	6.146e-05
ISRAEL	1.422e-06	1.336e-06	1.404e-05	STAIR	5.549e-04	8.566e-06	2.861e-05
STANDATA	5.645e-08	2.735e-07	5.130e-06	STANDGUB	2.934e-08	1.467e-07	2.753e-06
STOCFOR1	6.633e-09	9.701e-09	4.811e-08	VTP-BASE	1.349e-10	5.098e-11	2.342e-10

Table 1: Solving NETLIB LPs in **1000** iterations

4 Analysis

$$
\begin{aligned}
\nabla^{2}\varphi(\mathbf{x}) &= -\frac{\rho\nabla f(\mathbf{x})\nabla f(\mathbf{x})^{\top}}{f(\mathbf{x})^{2}}+\rho\frac{\mathbf{A}^{\top}\mathbf{A}}{f(\mathbf{x})}+\mathbf{X}^{-2} \\
\nabla\varphi(\mathbf{x}) &= \frac{\rho\nabla f(\mathbf{x})}{f(\mathbf{x})}-\mathbf{X}^{-1}\mathbf{e}
\end{aligned}
$$

$$
\begin{pmatrix} \mathbf{0}_{m\times m} & \mathbf{A} & \mathbf{0}_{m\times n} & \mathbf{0}_{m\times 1} & -\mathbf{b} \\ -\mathbf{A}^\top & \mathbf{0}_{n\times n} & -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times 1} & \mathbf{c} \\ \mathbf{b}^\top & -\mathbf{c}^\top & \mathbf{0}_{1\times n} & -1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{s} \\ \kappa \\ \tau \end{pmatrix} = \mathbf{0}
$$

$$
\mathbf{e}_n^\top \mathbf{x} + \mathbf{e}_n^\top \mathbf{s} + \kappa + \tau = 1.
$$

$$
\nabla^2 \varphi(\mathbf{x}) = -\frac{4\rho \mathbf{A}^\top \mathbf{A}\mathbf{x}\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{x}\|^4} + \frac{2\rho \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{x}\|^2} + \mathbf{X}^{-2}
$$

$$
\mathbf{X}\nabla^2 \varphi(\mathbf{x})\mathbf{X} = -\frac{4\rho \mathbf{X}\mathbf{A}^\top \mathbf{A}\mathbf{x}\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{X}}{\|\mathbf{A}\mathbf{x}\|^4} + \frac{2\rho \mathbf{X}\mathbf{A}^\top \mathbf{A}\mathbf{X}}{\|\mathbf{A}\mathbf{x}\|^2} + \mathbf{I}
$$

5 General Potential Method

$$
\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{e}^\top \mathbf{x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}
$$

$$
\begin{aligned}
\phi(\mathbf{x}) &= \log(f(\mathbf{x})) + \sum_{i=1}^n \log x_i \\
\nabla \phi(\mathbf{x}) &= \frac{\rho \nabla f(\mathbf{x})}{f(\mathbf{x})} - \mathbf{X}^{-1}\mathbf{e} \\
\nabla^2 \phi(\mathbf{x}) &= -\frac{\rho \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^\top}{f(\mathbf{x})^2} + \rho \frac{\nabla^2 f(\mathbf{x})}{f(\mathbf{x})} + \mathbf{X}^{-2} \\
f(\mathbf{x}) &\leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{L_1}{2}\|\mathbf{x}-\mathbf{y}\|^2 \\
f(\mathbf{x}) &\leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{1}{2}\left\langle (\mathbf{x}-\mathbf{y}) \nabla^2 f(\mathbf{y}), \mathbf{x}-\mathbf{y}\right\rangle + \frac{L_2}{6}\|\mathbf{x}-\mathbf{y}\|^3 \\
f(\mathbf{x}) &\geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle
\end{aligned}
$$

5.1 Second-order Potential Reduction

In this section, we consider the second order potential reduction method, where we update the iterates by

$$
\begin{aligned}
\mathbf{d}^k &= \mathop{\arg\min}_{\|\mathbf{d}\|\leq \beta, \mathbf{x}^\top \mathbf{d}=0} \left\{ \left\langle \mathbf{X}^k \nabla \phi\left(\mathbf{x}^k\right), \mathbf{d}\right\rangle + \frac{1}{2}\left\langle \mathbf{d}, \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}\right\rangle \right\} \\
\mathbf{x}^{k+1} &= \mathbf{x}^k + \mathbf{X}^k \mathbf{d}^k
\end{aligned}
$$

First, by the optimality condition of the trust region subproblem, we have, for some $\lambda^k \geq 0, \mu^k$ that

$$\begin{aligned}
\left(\mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k + \lambda^k \mathbf{I}\right) \mathbf{d}^k - \mu \mathbf{x}^k &= -\mathbf{X}^k \nabla \phi\left(\mathbf{x}^k\right) \\
\mu^k\left(\left\|\mathbf{d}^k\right\| - \beta\right) &= 0 \\
\mathbf{x}^\top \mathbf{d}^k &= 0 \\
\mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k + \lambda^k \mathbf{I} &\succeq_{\mathbf{x}} \mathbf{0}
\end{aligned}$$

Assume that $\left\|\mathbf{d}^k\right\| = \beta$ and define

$$p\left(\mathbf{x}, \mu\right) \quad := \quad \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k + \mathbf{X}^k \nabla \phi\left(\mathbf{x}^k\right) - \mu^k \mathbf{x}^k.$$

Then it follows that

$$\lambda^k \mathbf{d}^k \quad = \quad -p\left(\mathbf{x}^k, \mu^k\right)$$

and we successively deduce that

$$\begin{aligned}
& \left\langle \mathbf{X}^k \nabla \phi\left(\mathbf{x}^k\right), \mathbf{d}^k \right\rangle + \frac{1}{2}\left\langle \mathbf{d}^k, \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k \right\rangle \\
= \quad & \left\langle -\lambda^k \mathbf{d}^k - \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k + \mu^k \mathbf{x}^k, \mathbf{d}^k \right\rangle + \frac{1}{2}\left\langle \mathbf{d}^k, \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k \right\rangle \\
= \quad & -\lambda^k \left\|\mathbf{d}^k\right\|^2 - \frac{1}{2}\left\langle \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k, \mathbf{d}^k \right\rangle.
\end{aligned}$$

Since $\mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \succeq_{\mathbf{x}} -\lambda^k \mathbf{I}$, we have

$$\left\langle \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k, \mathbf{d}^k \right\rangle \geq -\left\|\mathbf{d}^k\right\|^2$$

and that

$$\left\langle \mathbf{X}^k \nabla \phi\left(\mathbf{x}^k\right), \mathbf{d}^k \right\rangle + \frac{1}{2}\left\langle \mathbf{d}^k, \mathbf{X}^k \nabla^2 \phi\left(\mathbf{x}^k\right) \mathbf{X}^k \mathbf{d}^k \right\rangle \leq -\frac{\lambda^k}{2}\left\|\mathbf{d}^k\right\|^2 = -\frac{\lambda^k \beta^2}{2}.$$

Next we derive the reduction of the potential function. It follows naturally that

$$\sum_{i=1}^n \log x_i - \sum_{i=1}^n \log(x_i + x_i d_i) \leq -\left\langle \mathbf{e}, \mathbf{d} \right\rangle + \frac{\beta^2}{2(1-\beta)}$$

First we bound the reduction in $\rho \log\left(f\left(\mathbf{x}\right)\right)$ by

$$\begin{aligned}
\log\left(\frac{f\left(\mathbf{x} + \mathbf{X}\mathbf{d}\right)}{f\left(\mathbf{x}\right)}\right) \quad &\leq \quad \log\left(1 + \frac{\left\langle \nabla f\left(\mathbf{x}\right), \mathbf{X}\mathbf{d} \right\rangle + \frac{L_1}{2}\left\|\mathbf{X}\mathbf{d}\right\|^2}{f\left(\mathbf{x}\right)}\right) \\
&\leq \quad \frac{\left\langle \nabla f\left(\mathbf{x}\right), \mathbf{X}\mathbf{d} \right\rangle + \frac{L_1}{2}\left\|\mathbf{X}\mathbf{d}\right\|^2}{f\left(\mathbf{x}\right)} - \frac{1}{2}\left(\frac{\left\langle \nabla f\left(\mathbf{x}\right), \mathbf{X}\mathbf{d} \right\rangle + \frac{L_1}{6}\left\|\mathbf{X}\mathbf{d}\right\|^2}{f\left(\mathbf{x}\right)}\right)^2 \\
& \qquad + \frac{1}{3}\left(\frac{\left\langle \nabla f\left(\mathbf{x}\right), \mathbf{X}\mathbf{d} \right\rangle + \frac{L_1}{6}\left\|\mathbf{X}\mathbf{d}\right\|^2}{f\left(\mathbf{x}\right)}\right)^3,
\end{aligned}$$

or, alternatively,

$$\begin{aligned}
\rho \log\left(\frac{f(\mathbf{x}+\mathbf{X}\mathbf{d})}{f(\mathbf{x})}\right) &\leq \rho \log\left(1+\frac{\langle \nabla f(\mathbf{x}), \mathbf{X}\mathbf{d}\rangle + \frac{1}{2}\left\langle \mathbf{d}^\top \mathbf{X}\nabla^2 f(\mathbf{x}), \mathbf{X}\mathbf{d}\right\rangle + \frac{L_2}{6}\|\mathbf{X}\mathbf{d}\|^3}{f(\mathbf{x})}\right) \\
&\leq \rho \frac{\langle \nabla f(\mathbf{x}), \mathbf{X}\mathbf{d}\rangle + \frac{1}{2}\left\langle \mathbf{d}^\top \mathbf{X}\nabla^2 f(\mathbf{x}), \mathbf{X}\mathbf{d}\right\rangle + \frac{L_2}{6}\|\mathbf{X}\mathbf{d}\|^3}{f(\mathbf{x})} \\
&= \langle \mathbf{X}\nabla\phi(\mathbf{x}), \mathbf{d}\rangle + \frac{1}{2}\left\langle \mathbf{d}^\top \mathbf{X}\nabla^2\phi(\mathbf{x})\mathbf{X}, \mathbf{d}\right\rangle \\
&\quad + \langle \mathbf{e}, \mathbf{d}\rangle - \|\mathbf{d}\|^2 + \frac{\rho \mathbf{d}^\top \mathbf{X}\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top \mathbf{X}\mathbf{d}}{2 f(\mathbf{x})^2} + \frac{\rho L_2}{6 f(\mathbf{x})}\|\mathbf{X}\mathbf{d}\|^3.
\end{aligned}$$

and

$$\begin{aligned}
& \phi(\mathbf{x}+\mathbf{X}\mathbf{d}) - \phi(\mathbf{x}) \\
= \; & \rho \log\left(\frac{f(\mathbf{x}+\mathbf{X}\mathbf{d})}{f(\mathbf{x})}\right) + \sum_{i=1}^n \log x_i - \sum_{i=1}^n \log(x_i + x_i d_i) \\
= \; & \langle \mathbf{X}\nabla\phi(\mathbf{x}), \mathbf{d}\rangle + \frac{1}{2}\left\langle \mathbf{d}^\top \mathbf{X}\nabla^2\phi(\mathbf{x})\mathbf{X}, \mathbf{d}\right\rangle \\
& + \langle \mathbf{e}, \mathbf{d}\rangle - \|\mathbf{d}\|^2 - \langle \mathbf{e}, \mathbf{d}\rangle + \frac{\beta^2}{2(1-\beta)} + \frac{\rho \mathbf{d}^\top \mathbf{X}\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top \mathbf{X}\mathbf{d}}{2 f(\mathbf{x})^2} \\
\leq \; & -\frac{\lambda}{2}\|\mathbf{d}\|^2 - \|\mathbf{d}\|^2 + \frac{\beta^2}{2(1-\beta)} + \frac{2\rho\gamma\beta^2}{n}
\end{aligned}$$