Interior Point Algorithms II

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Recall Path-Following Algorithms for LP

In general, one can start from an (approximate) central path point $\mathbf{x}(\mu^0)$, $(\mathbf{y}(\mu^0), \mathbf{s}(\mu^0))$, or $(\mathbf{x}(\mu^0), \mathbf{y}(\mu^0), \mathbf{s}(\mu^0))$ where μ^0 is sufficiently large.

Then, let μ^1 be a slightly smaller parameter than μ^0. Then, we compute an (approximate) central path point $\mathbf{x}(\mu^1)$, $(\mathbf{y}(\mu^1), \mathbf{s}(\mu^1))$, or $(\mathbf{x}(\mu^1), \mathbf{y}(\mu^1), \mathbf{s}(\mu^1))$. They can be updated from the previous point at μ^0 using the Newton method.

μ might be reduced at each stage by a specific factor, giving $\mu^{k+1} = \gamma\mu^k$ where γ is a fixed positive constant less than one, and k is the stage count.

This is called the primal, dual, or primal-dual path-following method.

LP Primal-Dual Path-Following Algorithm I

More precisely, given a pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}$ and

$$\|\mu\mathbf{e} - XS\mathbf{e}\| \le \frac{1}{3}\mu, \quad \text{where } \mu = \frac{\mathbf{x}^T\mathbf{s}}{n},$$

we can compute direction vectors $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the Newton system equations:

$$\begin{aligned} S\mathbf{d}_x + X\mathbf{d}_s &= (1 - \tfrac{1}{3\sqrt{n}})\mu\mathbf{e} - XS\mathbf{e}, \\ A\mathbf{d}_x &= \mathbf{0}, \\ -A^T\mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{1}$$

Note that $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T \mathbf{d}_y = 0$.

LP Primal-Dual Path-Following Algorithm II

Let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x,\ \mathbf{y}^+ = \mathbf{y} + \mathbf{d}_y,\ \mathbf{s}^+ = \mathbf{s} + \mathbf{d}_s$. Then, we have

Theorem 1

$$(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \mathrm{int}\,\mathcal{F},$$

and

$$\|\mu^+\mathbf{e} - X^+S^+\mathbf{e}\| \le \frac{1}{3}\mu^+, \quad \textit{where } \mu^+ = \frac{(\mathbf{x}^+)^T\mathbf{s}^+}{n} = (1 - \frac{1}{3\sqrt{n}})\mu.$$

It is easy to see

$$A\mathbf{x}^+ = \mathbf{b} \quad \text{and} \quad A^T\mathbf{y}^+ + \mathbf{s}^+ = \mathbf{c}.$$

Proof Sketch

$$X^{-0.5}S^{0.5}\mathbf{d}_x + S^{-0.5}X^{0.5}\mathbf{d}_s = (XS)^{-0.5}\left(\mu\mathbf{e} - XS\mathbf{e} - \frac{\mu}{3\sqrt{n}}\mathbf{e}\right).$$

$$\begin{aligned}
\|X^{-0.5}S^{0.5}\mathbf{d}_x + S^{-0.5}X^{0.5}\mathbf{d}_s\|^2 &\le \|(XS)^{-1}\| \cdot \|\left(\mu\mathbf{e} - XS\mathbf{e} - \tfrac{\mu}{3\sqrt{n}}\mathbf{e}\right)\|^2 \\
&= \tfrac{1}{\min(XS\mathbf{e})} \cdot \left(\|\mu\mathbf{e} - XS\mathbf{e}\|^2 + \tfrac{\mu^2}{9n}\|\mathbf{e}\|^2\right) \\
&\le \tfrac{3}{2\mu}\left(\tfrac{\mu^2}{9} + \tfrac{\mu^2}{9}\right) = \tfrac{\mu}{3}.
\end{aligned}$$

$$\|X^{-0.5}S^{0.5}\mathbf{d}_x\|^2 + \|S^{-0.5}X^{0.5}\mathbf{d}_s\|^2 = \|X^{-0.5}S^{0.5}\mathbf{d}_x + S^{-0.5}X^{0.5}\mathbf{d}_s\|^2 \le \frac{\mu}{3}.$$

$$\|S^{-1}\mathbf{d}_x\|^2 + \|X^{-1}\mathbf{d}_s\|^2 = \|(XS)^{-0.5}X^{0.5}S^{-0.5}\mathbf{d}_x\|^2 + \|(XS)^{-0.5}X^{-0.5}S^{0.5}\mathbf{d}_s\|^2$$

$$\le \|(XS)^{-1}\|(\|X^{-0.5}S^{0.5}\mathbf{d}_x\|^2 + \|X^{0.5}S^{-0.5}\mathbf{d}_s\|^2) \le \frac{1}{\min(XS\mathbf{e})}\cdot\frac{\mu}{3} \le \frac{3}{2\mu}\cdot\frac{\mu}{3} = \frac{1}{2}.$$

Proof Sketch continued

Summing the first set of equations:

$$(\mathbf{x}^+)^T\mathbf{s}^+ = (\mathbf{x}+\mathbf{d}_x)^T(\mathbf{s}+\mathbf{d}_s) = (1-\frac{1}{3\sqrt{n}})\mu.$$

$$\begin{aligned}
\|\mu^+\mathbf{e} - X^+S^+\mathbf{e}\| &= \|\mu^+ - XS\mathbf{e} - S\mathbf{d}_x - X\mathbf{d}_s - D_xD_s\mathbf{e}\| \\
&= \|D_xD_s\mathbf{e}\| = \|(X^{-0.5}S^{0.5}D_x)S^{-0.5}X^{0.5}D_s\mathbf{e}\| \\
&\le \tfrac{1}{2}(\|X^{-0.5}S^{0.5}\mathbf{d}_x + S^{-0.5}X^{0.5}\mathbf{d}_s\|^2) \\
&\le \tfrac{1}{2}\tfrac{\mu}{3} \le \tfrac{1}{3}\mu^+.
\end{aligned}$$

The first inequality of the above is from that, for any $\mathbf{a}\in R^n$ and $\mathbf{b}\in R^n$ with $\mathbf{a}^T\mathbf{b}=0$, we have

$$\sqrt{\sum_{j=1}^n |a_jb_j|^2} \le \sum_{j=1}^n |a_jb_j| \le \frac{1}{2}\left(\sum_{j=1}^n (a_j)^2 + \sum_{j=1}^n (b_j)^2\right).$$

Primal-Dual Potential Function for LP

For $\mathbf{x} \in \operatorname{int} \mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d$, the joint primal-dual potential function is defined by

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n + \rho) \log(\mathbf{x}^T \mathbf{s}) - \sum_{j=1}^{n} \log(x_j s_j),$$

where $\rho \geq 0$.

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) = \rho \log(\mathbf{x}^T \mathbf{s}) + \psi_n(\mathbf{x}, \mathbf{s}) \geq \rho \log(\mathbf{x}^T \mathbf{s}) + n \log n,$$

then, for $\rho > 0$, $\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \to -\infty$ implies that $\mathbf{x}^T \mathbf{s} \to 0$. More precisely, we have

$$\mathbf{x}^T \mathbf{s} \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) - n \log n}{\rho}\right).$$

Primal-Dual Potential Reduction Algorithm for LP

Once have a pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathrm{int}\, \mathcal{F}$, we again compute direction vectors $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the Newton system equations:

$$\begin{aligned} S\mathbf{d}_x + X\mathbf{d}_s &= \frac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - XS\mathbf{e}, \\ A\mathbf{d}_x &= \mathbf{0}, \\ -A^T\mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{2}$$

Note that $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T\mathbf{d}_y = 0$ here.

To simplify rotations, let

$$\begin{aligned} \mathbf{d}_{x'} + \mathbf{d}_{s'} &= \mathbf{r}' := (XS)^{-0.5}(\frac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - XS\mathbf{e}), \\ A'\mathbf{d}_{x'} &= \mathbf{0}, \\ -(A')^T\mathbf{d}_y - \mathbf{d}_{s'} &= \mathbf{0}. \end{aligned}$$

where

$$D = X^{0.5}S^{-0.5},\ A' = AD,\ \mathbf{d}_{x'} = D^{-1}\mathbf{d}_x,\ \mathbf{d}_{s'} = D\mathbf{d}_s.$$

Again, we maintain $\mathbf{d}_{x'}^T\mathbf{d}_{s'} = 0$.

Unlike in the path-following algorithm, $\|\mathbf{r}'\|^2$ may be too big to make $\mathbf{x} + \mathbf{d}_x$ or $\mathbf{s} + \mathbf{d}_s$ positive. So that we need to add a step size θ to scale $\mathbf{r}'$ such that it makes new iterate feasible.

Lemma 1 *Let the direction vector* $\mathbf{d} = (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ *be generated by equation (2), and let*

$$\theta = \frac{\alpha\sqrt{\min(XS\mathbf{e})}}{\|\mathbf{r}'\|}, \tag{3}$$

where α *is a positive constant less than 1. Let*

$$\mathbf{x}^+ = \mathbf{x} + \theta\mathbf{d}_x, \quad \mathbf{y}^+ = \mathbf{y} + \theta\mathbf{d}_y, \quad \textit{and} \quad \mathbf{s}^+ = \mathbf{s} + \theta\mathbf{d}_s.$$

Then, we have $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \operatorname{int}\mathcal{F}$ *and*

$$\psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s})$$

$$\leq -\alpha\sqrt{\min(XS\mathbf{e})}\|(XS)^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{x}^T\mathbf{s}}X\mathbf{s})\| + \frac{\alpha^2}{2(1-\alpha)}.$$

Logarithmic Approximation Lemma

We first present a technical lemma:

Lemma 2 *If* $\mathbf{d} \in \mathcal{R}^n$ *such that* $\|\mathbf{d}\|_\infty < 1$ *then*

$$\mathbf{e}^T\mathbf{d} \geq \sum_{i=1}^{n} \log(1 + d_i) \geq \mathbf{e}^T\mathbf{d} - \frac{\|\mathbf{d}\|^2}{2(1 - \|\mathbf{d}\|_\infty)}.$$

The proof is based on the Taylor expansion of $\ln(1 + d_i)$ for $-1 < d_i < 1$.

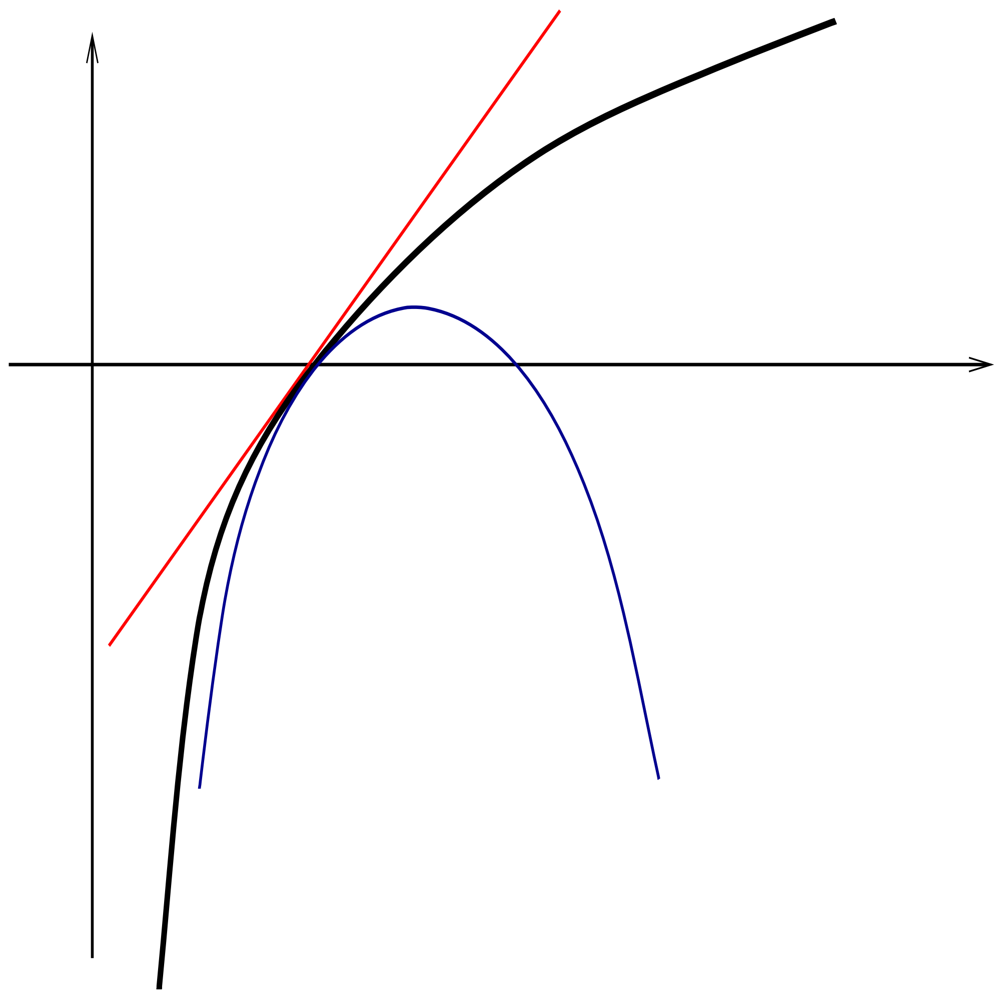

Figure 1: Logarithmic approximation by linear and quadratic functions

Proof Sketch of the Theorem

It is clear that $A\mathbf{x}^+ = \mathbf{b}$ and $A^T\mathbf{y}^+ + \mathbf{s}^+ = \mathbf{c}$. We now show that $\mathbf{x}^+ > \mathbf{0}$ and $\mathbf{s}^+ > \mathbf{0}$. This is similar to the previous proof for the path-following algorithm

$$\|\theta X^{-1}\mathbf{d}_x\|^2 + \|\theta S^{-1}\mathbf{d}_s\|^2 \le \theta^2 \frac{\|\mathbf{r}'\|^2}{\min(XS\mathbf{e})} = \frac{\alpha^2 \min(XS\mathbf{e})}{\|\mathbf{r}'\|^2} \frac{\|\mathbf{r}'\|^2}{\min(XS\mathbf{e})} = \alpha^2 < 1.$$

Therefore,

$$\mathbf{x}^+ = \mathbf{x} + \theta\mathbf{d}_x = X(\mathbf{e} - \theta X^{-1}\mathbf{d}_x) > \mathbf{0}$$

and

$$\mathbf{s}^+ = \mathbf{s} + \theta\mathbf{d}_s = S(\mathbf{e} - \theta S^{-1}\mathbf{d}_s) > \mathbf{0}.$$

Sketch of the proof continued

$$
\begin{aligned}
& \psi(\mathbf{x}^+, \mathbf{s}^+) - \psi(\mathbf{x}, \mathbf{s}) \\
= \;& (n+\rho)\log\left(1 + \frac{\theta \mathbf{d}_s^T\mathbf{x} + \theta \mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \sum_{j=1}^n \left(\log(1 + \frac{\theta d_{s_j}}{s_j}) + \log(1 + \frac{\theta d_{x_j}}{x_j})\right) \\
\le \;& (n+\rho)\left(\frac{\theta \mathbf{d}_s^T\mathbf{x} + \theta \mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \sum_{j=1}^n \left(\log(1 + \frac{\theta d_{s_j}}{s_j}) + \log(1 + \frac{\theta d_{x_j}}{x_j})\right) \\
\le \;& (n+\rho)\left(\frac{\theta \mathbf{d}_s^T\mathbf{x} + \theta \mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \theta\mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \frac{||\theta S^{-1}\mathbf{d}_s||^2 + ||\theta X^{-1}\mathbf{d}_x||^2}{2(1-\alpha)} \\
\le \;& \frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\theta(\mathbf{d}_s^T\mathbf{x} + \mathbf{d}_x^T\mathbf{s}) - \theta\mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= \;& \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\mathbf{e}^T(X\mathbf{d}_s + S\mathbf{d}_x) - \mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= \;& \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\mathbf{e}^T(X\mathbf{d}_s + S\mathbf{d}_x) - \mathbf{e}^T(XS)^{-1}(X\mathbf{d}_s + S\mathbf{d}_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= \;& \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}XS\mathbf{e} - \mathbf{e}\right)^T (XS)^{-1}(X\mathbf{d}_s + S\mathbf{d}_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= \;& \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}XS\mathbf{e} - \mathbf{e}\right)^T (XS)^{-1}\left(\frac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - XS\mathbf{e}\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= \;& -\theta\cdot\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\cdot\|\mathbf{r}'\|^2 + \frac{\alpha^2}{2(1-\alpha)} = -\alpha\sqrt{\min(XS\mathbf{e})}\cdot\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\cdot\|\mathbf{r}'\| + \frac{\alpha^2}{2(1-\alpha)}.
\end{aligned}
$$

Let $\mathbf{v} = XS\mathbf{e}$. Then, we can prove the following technical lemma:

Lemma 3 *Let $\mathbf{v} \in \mathcal{R}^n$ be a positive vector and $\rho \geq \sqrt{n}$. Then,*

$$\sqrt{\min(\mathbf{v})}\|V^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{e}^T\mathbf{v}}\mathbf{v})\| \geq \sqrt{3/4}\,.$$

Combining these two lemmas we have

$$\psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s})$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)} = -\delta$$

for a constant δ.

Description of Algorithm

Given $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0) \in \operatorname{int} \mathcal{F}$. Set $\rho \geq \sqrt{n}$ and $k := 0$.

While $(\mathbf{x}^k)^T \mathbf{s}^k \geq \epsilon$ **do**

1. Set $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$ and $\gamma = n/(n+\rho)$ and compute $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from (2).
2. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + \bar{\alpha}\mathbf{d}_x$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \bar{\alpha}\mathbf{d}_y$, and $\mathbf{s}^{k+1} = \mathbf{s}^k + \bar{\alpha}\mathbf{d}_s$ where
$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(\mathbf{x}^k + \alpha\mathbf{d}_x, \mathbf{s}^k + \alpha\mathbf{d}_s).$$
3. Let $k := k + 1$ and return to Step 1.

Theorem 2 *Let $\rho \geq \sqrt{n}$ and $\psi_{n+\rho}(\mathbf{x}^0, \mathbf{s}^0) \leq \rho \log((\mathbf{x}^0)^T \mathbf{s}^0) + n \log n$. Then, the Algorithm terminates in at most $O(\rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon))$ iterations with*

$$(\mathbf{x}^k)^T \mathbf{s}^k = \mathbf{c}^T \mathbf{x}^k - \mathbf{b}^T \mathbf{y}^k \leq \epsilon.$$

$$\begin{aligned}
(\mathbf{x}^k)^T \mathbf{s}^k \quad & \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^k,\mathbf{s}^k) - n \log n}{\rho}\right) \\
& \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^0,\mathbf{s}^0) - n \log n - \rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon)}{\rho}\right) \\
& \leq \exp\left(\frac{\rho \log(\mathbf{x}^0,\mathbf{s}^0) - \rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon)}{\rho}\right) \\
& = \exp(\log(\epsilon)) = \epsilon.
\end{aligned}$$

The role of ρ? And aggressive step size?

Alternating Direction Method

Recall that for $\mathbf{x} \in \text{int}\,\mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}_d$, the joint primal-dual potential function is defined as

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n+\rho)\log(\mathbf{x}^T\mathbf{s}) - \sum_{j=1}^{n}\log(x_j s_j)$$

$$= (n+\rho)\log(\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}) - \sum_{j=1}^{n}\log(x_j) - \sum_{j=1}^{n}\log(s_j).$$

The algorithm we described earlier is a simultaneous updating on both primal and dual.

One can also develop a method by alternate updating primal $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s})$. More precisely,, at the kth step, fix $(\mathbf{y}^k, \mathbf{s}^k)$ and reduce the potential function by a constant via updating from $\mathbf{x}^k$ to $\mathbf{x}^{k+1}$ while keep $(\mathbf{y}^{k+1}, \mathbf{s}^{k+1}) = (\mathbf{y}^k, \mathbf{s}^k)$:

$$\psi_{n+\rho}(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) - \psi_{n+\rho}(\mathbf{x}^k, \mathbf{s}^k) \leq -\delta.$$

Once can prove that, if by updating primal only one cannot reduce the potential function by a constant anymore, then one must be able to update the dual from $(\mathbf{y}^k, \mathbf{s}^k)$ to $(\mathbf{y}^{k+1}, \mathbf{s}^{k+1})$ (while keep $\mathbf{x}^{k+1} = \mathbf{x}^k$) and reduce the potential function by a constant. Thus, the sample complexity result holds.