A

Convergence Proof

recall $r^k = Ax^k + Bz^k - c$

The basic convergence result given in §3.2 can be found in several references, such as [81, 63]. Many of these give more sophisticated results, with more general penalties or inexact minimization. For completeness, we give a proof here.

We will show that if f and g are closed, proper, and convex, and the Lagrangian L_0 has a saddle point, then we have primal residual convergence, meaning that $r^k \to 0$, and objective convergence, meaning that $p^k \to p^\star$, where $\boxed{p^k = f(x^k) + g(z^k).}$ We will also see that the dual residual $s^k = \rho A^T B(z^k - z^{k-1})$ converges to zero.

Let $(x^\star, z^\star, y^\star)$ be a saddle point for L_0, and define

$$\boxed{V^k = (1/\rho)\|y^k - y^\star\|_2^2 + \rho\|B(z^k - z^\star)\|_2^2,}$$

We will see that V^k is a *Lyapunov function* for the algorithm, *i.e.*, a nonnegative quantity that decreases in each iteration. (Note that V^k is unknown while the algorithm runs, since it depends on the unknown values $z^\star$ and $y^\star$.)

We first outline the main idea. The proof relies on three key inequalities, which we will prove below using basic results from convex analysis along with simple algebra. The first inequality is

$$V^{k+1} \leq V^k - \rho\|r^{k+1}\|_2^2 - \rho\|B(z^{k+1} - z^k)\|_2^2. \tag{A.1}$$

This states that V^k decreases in each iteration by an amount that depends on the norm of the residual and on the change in z over one iteration. Because $V^k \leq V^0$, it follows that y^k and Bz^k are bounded. Iterating the inequality above gives that

$$\rho \sum_{k=0}^{\infty} \left(\|r^{k+1}\|_2^2 + \|B(z^{k+1} - z^k)\|_2^2 \right) \leq V^0,$$

which implies that $r^k \to 0$ and $B(z^{k+1} - z^k) \to 0$ as $k \to \infty$. Multiplying the second expression by ρA^T shows that the dual residual $s^k = \rho A^T B(z^{k+1} - z^k)$ converges to zero. (This shows that the stopping criterion (3.12), which requires the primal and dual residuals to be small, will eventually hold.)

The second key inequality is

$$\begin{aligned} &p^{k+1} - p^\star \\ &\quad \leq -(y^{k+1})^T r^{k+1} - \rho(B(z^{k+1} - z^k))^T(-r^{k+1} + B(z^{k+1} - z^\star)), \end{aligned} \tag{A.2}$$

and the third inequality is

$$p^\star - p^{k+1} \leq y^{\star T} r^{k+1}. \tag{A.3}$$

The righthand side in (A.2) goes to zero as $k \to \infty$, because $B(z^{k+1} - z^\star)$ is bounded and both r^{k+1} and $B(z^{k+1} - z^k)$ go to zero. The righthand side in (A.3) goes to zero as $k \to \infty$, since r^k goes to zero. Thus we have $\lim_{k\to\infty} p^k = p^\star$, *i.e.*, objective convergence.

Before giving the proofs of the three key inequalities, we derive the inequality (3.11) mentioned in our discussion of stopping criterion from the inequality (A.2). We simply observe that $-r^{k+1} + B(z^{k+1} - z^k) = -A(x^{k+1} - x^\star)$; substituting this into (A.2) yields (3.11),

TYPOS SHOULD BE $z^\star$

$$p^{k+1} - p^\star \leq -(y^{k+1})^T r^{k+1} + (x^{k+1} - x^\star)^T s^{k+1}.$$

$$r^{k+1} = Ax^{k+1} + Bz^{k+1} - \underbrace{(Ax^\star + Bz^\star)}_{c}$$

Proof of inequality (A.3)

Since $(x^\star, z^\star, y^\star)$ is a saddle point for L_0, we have

$$L_0(x^\star, z^\star, y^\star) \leq L_0(x^{k+1}, z^{k+1}, y^\star).$$

$$p^{k+1} + (y^{k+1})^T(Ax^{k+1} + Bz^{k+1}) - \rho(B(z^{k+1} - z^k))^T Ax^{k+1}$$
$$\le p^* + (y^{k+1})^T(\underbrace{Ax^* + Bz^*}_{=c}) - \rho(B(z^{k+1} - z^k))^T Ax^*$$
$$p^{k+1} - p^* \le (-y^{k+1})^T r^{k+1} - \rho(B(z^{k+1} - z^k))^T \underbrace{A(x^* - x^{k+1})}$$

$$(Ax^* + Bz^*) - (Ax^{k+1} + Bz^{k+1} + B(z^* - z^{k+1}))$$

Using $Ax^\star + Bz^\star = c$, the lefthand side is $p^\star$. With $p^{k+1} = f(x^{k+1}) + g(z^{k+1})$, this can be written as

$$\boxed{p^\star \le p^{k+1} + y^{\star T} r^{k+1},}$$

which gives (A.3).

Proof of inequality (A.2)

$$= f(x) + g(z^k) + (y^k)^T(Ax + Bz^k - c) + \frac{\rho}{2}\|Ax + Bz^k - c\|^2$$

By definition, x^{k+1} minimizes $L_\rho(x, z^k, y^k)$. Since f is closed, proper, and convex it is subdifferentiable, and so is L_ρ. The (necessary and sufficient) optimality condition is

$$\boxed{0 \in \partial L_\rho(x^{k+1}, z^k, y^k) = \partial f(x^{k+1}) + A^T y^k + \rho A^T(Ax^{k+1} + Bz^k - c).}$$

(Here we use the basic fact that the subdifferential of the sum of a subdifferentiable function and a differentiable function with domain $\mathbf{R}^n$ is the sum of the subdifferential and the gradient; see, *e.g.*, [140, §23].)

Since $y^{k+1} = y^k + \rho r^{k+1}$, we can plug in $y^k = y^{k+1} - \rho r^{k+1}$ and rearrange to obtain

$$\boxed{0 \in \partial f(x^{k+1}) + A^T(y^{k+1} - \rho B(z^{k+1} - z^k)).}$$

This implies that x^{k+1} minimizes

$$f(x) + (y^{k+1} - \rho B(z^{k+1} - z^k))^T Ax.$$

A similar argument shows that z^{k+1} minimizes $g(z) + y^{(k+1)T} Bz$. It follows that

$$\begin{aligned} f(x^{k+1}) &+ (y^{k+1} - \rho B(z^{k+1} - z^k))^T Ax^{k+1} \\ &\le f(x^\star) + (y^{k+1} - \rho B(z^{k+1} - z^k))^T Ax^\star \end{aligned}$$

and that

$$g(z^{k+1}) + y^{(k+1)T} Bz^{k+1} \le g(z^\star) + y^{(k+1)T} Bz^\star.$$

Adding the two inequalities above, using $Ax^\star + Bz^\star = c$, and rearranging, we obtain (A.2).

as z^{k+1} minimizes $L_\rho(x^{k+1}, z, y^k) = f(x^{k+1}) + g(z) + (y^k)^T(Ax^{k+1} + Bz - c) + \frac{\rho}{2}\| \nearrow \|^2$

so $0 \in \partial g(z^{k+1}) + B^T y^k + \frac{\rho}{2} B^T r^{k+1}$

$0 \in \partial g(z^{k+1}) + B^T y^{k+1}$

Proof of inequality (A.1) (omitted)

Adding (A.2) and (A.3), regrouping terms, and multiplying through by 2 gives

$$\begin{aligned}2(y^{k+1}-y^\star)^T r^{k+1} &- 2\rho(B(z^{k+1}-z^k))^T r^{k+1}\\ &+ 2\rho(B(z^{k+1}-z^k))^T(B(z^{k+1}-z^\star)) \le 0.\end{aligned} \tag{A.4}$$

The result (A.1) will follow from this inequality after some manipulation and rewriting.

We begin by rewriting the first term. Substituting $y^{k+1} = y^k + \rho r^{k+1}$ gives

$$2(y^k - y^\star)^T r^{k+1} + \rho\|r^{k+1}\|_2^2 + \rho\|r^{k+1}\|_2^2,$$

and substituting $r^{k+1} = (1/\rho)(y^{k+1} - y^k)$ in the first two terms gives

$$(2/\rho)(y^k - y^\star)^T(y^{k+1} - y^k) + (1/\rho)\|y^{k+1} - y^k\|_2^2 + \rho\|r^{k+1}\|_2^2.$$

Since $y^{k+1} - y^k = (y^{k+1} - y^\star) - (y^k - y^\star)$, this can be written as

$$(1/\rho)\left(\|y^{k+1} - y^\star\|_2^2 - \|y^k - y^\star\|_2^2\right) + \rho\|r^{k+1}\|_2^2. \tag{A.5}$$

We now rewrite the remaining terms, *i.e.*,

$$\rho\|r^{k+1}\|_2^2 - 2\rho(B(z^{k+1}-z^k))^T r^{k+1} + 2\rho(B(z^{k+1}-z^k))^T(B(z^{k+1}-z^\star)),$$

where $\rho\|r^{k+1}\|_2^2$ is taken from (A.5). Substituting

$$z^{k+1} - z^\star = (z^{k+1} - z^k) + (z^k - z^\star)$$

in the last term gives

$$\begin{aligned}\rho\|r^{k+1} - B(z^{k+1}-z^k)\|_2^2 &+ \rho\|B(z^{k+1}-z^k)\|_2^2\\ &+2\rho(B(z^{k+1}-z^k))^T(B(z^k - z^\star)),\end{aligned}$$

and substituting

$$z^{k+1} - z^k = (z^{k+1} - z^\star) - (z^k - z^\star)$$

in the last two terms, we get

$$\rho\|r^{k+1} - B(z^{k+1}-z^k)\|_2^2 + \rho\left(\|B(z^{k+1}-z^\star)\|_2^2 - \|B(z^k - z^\star)\|_2^2\right).$$

With the previous step, this implies that (A.4) can be written as

$$V^k - V^{k+1} \geq \rho \|r^{k+1} - B(z^{k+1} - z^k)\|_2^2. \tag{A.6}$$

To show (A.1), it now suffices to show that the middle term $-2\rho r^{(k+1)T}(B(z^{k+1} - z^k))$ of the expanded right hand side of (A.6) is positive. To see this, recall that z^{k+1} minimizes $g(z) + y^{(k+1)T}Bz$ and z^k minimizes $g(z) + y^{kT}Bz$, so we can add

$$g(z^{k+1}) + y^{(k+1)T}Bz^{k+1} \leq g(z^k) + y^{(k+1)T}Bz^k$$

and

$$g(z^k) + y^{kT}Bz^k \leq g(z^{k+1}) + y^{kT}Bz^{k+1}$$

to get that

$$(y^{k+1} - y^k)^T(B(z^{k+1} - z^k)) \leq 0.$$

Substituting $y^{k+1} - y^k = \rho r^{k+1}$ gives the result, since $\rho > 0$.