# Provably Convergent Plug & Play Linearized ADMM, applied to Deblurring Spatially Varying Kernels

Charles Laroche, Andrés Almansa, Eva Coupeté, Matias Tassano

**Abstract**

Convergence proof.

## A Convergence of linearized-ADMM

In this section, we prove Theorem 1, without loss of generality we suppose that $\lambda = 1$ in our MAP estimator defined in ().

**Lemma 1.** *Under assumption 1, the following inequality holds for the $x$-update:*

$$\mathcal{L}_\beta(x_k, z_k, w_k) - \mathcal{L}_\beta(x_{k+1}, z_k, w_k) \geq \frac{L_x - \beta \|H\|^2}{2} \|x_k - x_{k+1}\|^2$$

*with $\|H\|^2$ the largest singular value of $H^T H$.*

*Proof.* Using the notation of **??**, we define $\overline{f}^k(x) = \tilde{\mathcal{L}}_\beta^k(x, z_k, w_k)$ and by definition of the $x$-update, we have:

$$\overline{f}^k(x_k) \geq \overline{f}^k(x_{k+1}) \tag{1}$$

$$\begin{aligned} \Leftrightarrow & \langle x_k - x_{k+1}, H^T w_k + \beta H^T (H x_k - z_k) \rangle \\ & + f(x_k) - f(x_{k+1}) \geq \frac{L_x}{2} \|x_{k+1} - x_k\|^2 \end{aligned} \tag{2}$$

We also have that:

$$
\begin{aligned}
&\mathcal{L}_\beta(x_k, z_k, w_k) - \mathcal{L}_\beta(x_{k+1}, z_k, w_k) \\
&= f(x_k) - f(x_{k+1}) + \langle w_k, H(x_k - x_{k+1}) \rangle \\
&\quad + \frac{\beta}{2}\|Hx_k - z_k\|^2 - \frac{\beta}{2}\|Hx_{k+1} - z_k\|^2 && (3) \\
&= f(x_k) - f(x_{k+1}) - \frac{\beta}{2}\|H(x_{k+1} - x_k)\|^2 \\
&\quad + \langle x_k - x_{k+1}, H^T w_k + \beta H^T(Hx_k - z_k) \rangle && (4) \\
&\geq \frac{L_x}{2}\|x_{k+1} - x_k\|^2 - \frac{\beta}{2}\|H(x_{k+1} - x_k)\|^2 && (5) \\
&\geq \frac{L_x - \beta\|H\|^2}{2}\|x_{k+1} - x_k\|^2 && (6)
\end{aligned}
$$

where the inequality (5) is obtained using (2). □

**Lemma 2.** $\mathcal{L}_\beta(x_{k+1}, z_k, w_k) - \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) \geq m\|z_k - z_{k+1}\|^2$

*Proof.* From Assumption 1, we have that $\mathcal{L}_\beta$ is strongly convex in z with parameter m. The strong convexity of $\mathcal{L}_\beta$ implies that:

$$
\begin{aligned}
&\mathcal{L}_\beta(x, z_k, u) - \mathcal{L}_\beta(x, z_{k+1}, u) && (7) \\
&\quad \geq \nabla_z \mathcal{L}_\beta(x, z_{k+1}, u)(z_k - z_{k+1}) + m\|z_k - z_{k+1}\|^2 && (8)
\end{aligned}
$$

However, the z-update of Algorithm 1 is such that

$$\nabla_z \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) = 0 \tag{9}$$

which leads to the results. □

**Lemma 3.** *Under Assumption 1, the following equality holds:*

$$w_k = \nabla_z h(z_k) \tag{10}$$

*Proof.* From the definition of the Lagrangian:

$$\nabla_z \mathcal{L}_\beta(x, z, w) = \nabla h(z) - w - \beta(Hx - z)$$

Using the fact that

$$w_{k+1} = w_k + \beta(Hx_{k+1} - z_{k+1}) \quad \text{and} \quad \nabla_z \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) = 0 \tag{11}$$

We have:

$$
\begin{aligned}
&0 = \nabla h(z_{k+1}) - w_k - \beta(Hx_{k+1} - z_{k+1}) && (12) \\
&\Leftrightarrow \nabla h(z_{k+1}) = w_{k+1} && (13)
\end{aligned}
$$

□

**Lemma 4.** *Under assumption 1,*

$$\mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1}) - \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) \tag{14}$$

$$= \frac{1}{\beta}\|w_{k+1} - w_k\|^2 \leq C_1\|z_{k+1} - z_k\|^2 \tag{15}$$

*with* $C_1 = L_h^2/\beta$.

*Proof.* By definition of the augmented Lagrangian we have that:

$$\begin{aligned}
\mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1}) &- \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) \\
&= \langle w_{k+1} - w_k, Hx_{k+1} - z_{k+1}\rangle \\
&= \frac{1}{\beta}\|w_{k+1} - w_k\|^2 \\
&= \frac{1}{\beta}\|\nabla_z h(z_{k+1}) - \nabla_z h(z_k)\|^2 \quad \text{from Lemma 3} \\
&\leq \frac{L_h^2}{\beta}\|z_{k+1} - z_k\|^2 \quad \text{from} \quad \text{Assumption 1}
\end{aligned}$$

□

**Lemma 5.** *Let* $g$ *is* $L_g$*-Lipschitz differentiable then:*

$$g(y_2) - g(y_1) \geq \nabla g(s)(y_2 - y_1) - \frac{L_g}{2}\|y_2 - y_1\|^2 \tag{16}$$

*where* $s$ *denotes* $y_1$ *or* $y_2$

*Proof.*

$$g(y_2) - g(y_1) = \int_0^1 \nabla g(ty_2 + (1-t)y_1) \cdot (y_2 - y_1)\mathrm{d}t \tag{17}$$

$$= \int_0^1 \nabla g(s) \cdot (y_2 - y_1)\mathrm{d}t + \int_0^1 (\nabla g(y_2 + (1-t)y_1) - \nabla g(s)) \cdot (y_2 - y_1)\mathrm{d}t, \tag{18}$$

where $\nabla g(\cdot)$ defines the gradient of $g(\cdot)$. If we take $s = y_1$, then by inequality

$$\|\nabla g(ty_2 + (1-t)y_1) - \nabla g(y_1)\| \leq L_g\|t(y_2 - y_1)\| \tag{19}$$

we have

$$\int_0^1 \nabla g(y_1) \cdot (y_2 - y_1)\mathrm{d}t + \int_0^1 (\nabla g(ty_2 + (1-t)y_1) - \nabla g(y_1)) \cdot (y_2 - y_1)\mathrm{d}t \tag{20}$$

$$\geq \nabla g(y_1) \cdot (y_2 - y_1) - \int_0^1 L_g t\|y_2 - y_1\|^2 \ \mathrm{d}t \tag{21}$$

$$= \nabla g(y_1) \cdot (y_2 - y_1) - \frac{L_g}{2}\|y_2 - y_1\|^2. \tag{22}$$

Therefore, we get

$$g(y_2) - g(y_1) \geq \nabla g(y_1) \cdot (y_2 - y_1) - \frac{L_g}{2} \|y_2 - y_1\|^2. \tag{23}$$

Similarly, if we take $s = y_2$, we can get

$$g(y_2) - g(y_1) \geq \nabla g(y_2) \cdot (y_2 - y_1) - \frac{L_g}{2} \|y_2 - y_1\|^2. \tag{24}$$

□

**Lemma 6.** *Under Assumption 1, if we choose the hyper-parameters $\beta$ and $L_x$ satisfying (**??**) and (**??**), then the sequence $\{m_k\}$ defined by*

$$m_k = \mathcal{L}_\beta(x_k, z_k, w_k) \tag{25}$$

*is convergent.*

*Proof.* 1) **Monotonicity:** By using Lemma 1, Lemma 2 and Lemma 4 we have:

$$\begin{aligned}
m_k - m_{k+1} &= \mathcal{L}_\beta(x_k, z_k, w_k) - \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1}) && (26)\\
&\geq \mathcal{L}_\beta(x_{k+1}, z_k, w_k) - \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1}) && (27)\\
&\quad + \frac{L_x - \beta\|H\|^2}{2} \|x_k - x_{k+1}\|^2 && (28)\\
&\geq \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_k) - \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1}) && (29)\\
&\quad + \frac{L_x - \beta\|H\|^2}{2} \|x_k - x_{k+1}\|^2 + m\|z_k - z_{k+1}\|^2 && (30)\\
&\geq \frac{L_x - \beta\|H\|^2}{2} \|x_k - x_{k+1}\|^2 + (m + \frac{L_h^2}{\beta})\|z_k - z_{k+1}\|^2 && (31)
\end{aligned}$$

Since we chose $L_x$ such that:

$$\begin{aligned}
& L_x \geq \beta\|H\|^2 && (32)\\
\Leftrightarrow \quad & \frac{L_x - \beta\|H\|^2}{2} > 0 && (33)
\end{aligned}$$

we obtain the monotonocity of $\{m_k\}$.

2) **Lower bound**:

$$m_k = h(z_k) + f(x_k) + \langle w_k, Hx_k - z_k \rangle + \frac{\beta}{2} \|Hx_k - z_k\|^2 \tag{34}$$

Let $z_k' = Hx_k$, from Lemma 3 we have:

$$\begin{aligned}
\langle w_k, Hx_k - z_k \rangle &= \langle w_k, z_k' - z_k \rangle && (35)\\
&= \langle \nabla h(z_k), z_k' - z_k \rangle && (36)
\end{aligned}$$

so we can rewrite:

$$m_k = h(z_k) + f(x_k) + \langle \nabla h(z_k), z'_k - z_k \rangle + \frac{\beta}{2}\|z'_k - z_k\|^2 \tag{37}$$

$$\tag{38}$$

We chose $\beta$ such that $\beta \geq L_h$ so:

$$\begin{aligned} m_k &\geq h(z_k) + f(x_k) - \langle \nabla h(z_k), z_k - z'_k \rangle + \frac{L_h}{2}\|z_k - z'_k\|^2 && (39) \\ &\geq h(z'_k) + f(x_k) \quad \text{from Lemma 5.} && (40) \end{aligned}$$

Following Assumption 1, $h(z'_k) + f(x_k)$ is lower bounded so $m_k$ is lower bounded. $m_k$ is monotonically decreasing and lower bounded which ensure the convergence. □

**Lemma 7.** *Suppose we have a differentiable function $f_1$, a possibly non differentiable function $f_2$, and a point $x$. If there exist $d2 \in \partial f_2(x)$, then we have:*

$$d = d_2 - \nabla f_1(x) \in \partial(f_2(x) - f_1(x))$$

*Proof.* From the subgradient definition we have that:

$$f_2(y) \geq f_2(x) + \langle d_2, y - x \rangle + o(\|y - x\|) \tag{41}$$

From the fact that $f_1$ is differentiable we have that:

$$-f_1(y) = -f_1(x) - \langle \nabla f_1(x), y - x \rangle + o(\|y - x\|) \tag{42}$$

Combining the two leads to:

$$f_2(y) - f_1(y) \geq f_2(x) - f_1(x) + \langle d_2 - \nabla f_1(x), y - x \rangle + o(\|y - x\|) \tag{43}$$

□

*Proof of Theorem 1:*
**Convergence of the residuals:**
From Lemma 6 and its proof we have that:

$$m_{k+1} - m_k \geq a\|x_k - x_{k+1}\|^2 + m\|z_{k-1} - z_k\|^2 \geq 0 \tag{44}$$

with $(m + \frac{L_h^2}{\beta}) > 0$, $a = \frac{L_x - \beta\|H\|^2}{2} > 0$ (according to Assumption 1) and that $m_k$ converges. This implies that $\|x_k - x_{k+1}\|^2$ and $\|y_k - y_{k+1}\|^2$ converge to 0 as k approaches infinity. Lemma 4 ensure the convergence of $\|w_k - w_{k+1}\|^2$ to 0. The convergence of $m_k$ directly implies the convergence of $\mathcal{L}_\beta(x_k, z_k, w_k)$.
**Convergence of the gradients:**
For the convergence of $\lim_{k\to\infty} \nabla_u \mathcal{L}_\beta(x_k, z_k, w_k)$, we have that:

$$\lim_{k\to\infty} \nabla_u \mathcal{L}_\beta(x_k, z_k, w_k) = \lim_{k\to\infty} Hx_k - z_k = \lim_{k\to\infty} \frac{1}{\beta}(w_{k+1} - w_k) = 0. \tag{45}$$

On the other side, we have using Lemma 3 that:

$$\nabla_z \mathcal{L}_\beta(x_k, z_k, w_k) = \nabla h(z_k) - w_k - \beta(Hx_k - z_k) \tag{46}$$
$$= w_k - w_k - (w_{k+1} - w_k) = -(w_{k+1} - w_k) \to 0 \tag{47}$$

Finally, we want to show that there exists

$$d^k \in \partial_x \mathcal{L}_\beta(x_k, z_k, w_k) \quad \text{s.t} \quad \lim_{k\to\infty} d^k = 0. \tag{48}$$

Since $x^{k+1}$ is the minimum point of $\tilde{\mathcal{L}}^k_\beta(x, z_k, w_k)$, we have that $0 \in \partial\tilde{\mathcal{L}}^k_\beta(x, z_k, w_k)$. Using Lemma 7 and the definition of $\tilde{\mathcal{L}}^k_\beta$ we have:

$$\exists d_{k+1} \in \partial f(x_{k+1}) \tag{49}$$
$$s.t \quad H^T w_k + L_x(x_{k+1} - x_k) + \beta H^T(Hx_k - z_k) + d_{k+1} = 0 \tag{50}$$

Lets us define:

$$\tilde{d}_{k+1} = H^T w_{k+1} + \beta H^T(Hx_{k+1} - z_{k+1}) + d_{k+1} \tag{51}$$

we can easily verify that $\tilde{d}_{k+1} \in \partial_x \mathcal{L}_\beta(x_{k+1}, z_{k+1}, w_{k+1})$.
We arleady showed that the primal residues $\|x_{k+1} - x_k\|$, $\|z_{k+1} - z_k\|$, $\|w_{k+1} - w_k\|$ converge to 0 as k approaches infinity, therefore:

$$\lim_{k\to\infty} \tilde{d}_{k+1} = \lim_{k\to\infty} H^T w_{k+1} + \beta H^T(Hx_{k+1} - z_{k+1}) + d_{k+1} \tag{52}$$
$$= \lim_{k\to\infty} H^T w_k + L_x(x_{k+1} - x_k) + \beta H^T(Hx_k - z_k) + d_{k+1} = 0 \tag{53}$$

where the last equality is obtained using 50.

# B Application to PnP-LADMM

## B.1 Proof of Proposition 1

### B.1.1 Proximal Gradient Step Denoiser.

*Proof.* Let $\mathcal{D}_{\sigma_d}$ be the proximal gradient step denoiser defined in [**?**] as $\mathcal{D}_{\sigma_d} := Id - \nabla g_{\sigma_d}$ where $g_{\sigma_d}(x) = \frac{1}{2}\|x - N_{\sigma_d}\|^2$ and $N_{\sigma_d}$ is a neural network.

According to [**?**, Proposition 3.1] there exists $\phi_{\sigma_d}$ such that $\mathcal{D}_{\sigma_d} = \mathrm{prox}_{\phi_{\sigma_d}}$.

In addition [**?**, Equation (26)] states that $\phi_{\sigma_d} \geq g_{\sigma_d}$, and by definition $g_{\sigma_d} \geq 0$.

Hence, $f = \phi_{\sigma_d}/\sigma_d^2$ is lower bounded by 0 and $\mathcal{D}_{\sigma_d} = \mathrm{prox}_{\sigma_d^2 f}$ as indicated by Proposition 1. $\square$

### B.1.2 MMSE Denoiser.

*Proof.* Let $\mathcal{D}_{\sigma_d}(y) = E[X|Y = y]$ be an MMSE denoiser, where $Y = X + \sigma_d N$ and $N \sim \mathcal{N}(0, \sigma_d^2 Id)$, and $X \sim p_X$, $p_X$ being a probability measure.

We want to show that there exists a lower bounded $\phi_{\sigma_d}$ such that $\mathcal{D}_{\sigma_d}(x) = \text{prox}_{\phi_{\sigma_d}}(x)$.

For $\sigma_d = 1$ according to [**?**] there exists $f(x) \geq -\log p_Y(x)$, such that $\mathcal{D}_1 = \text{prox}_f$. $f$ is lower bounded because the noisy density $p_Y(x) = (p_X * g_1)(x) \leq 1/\sqrt{2\pi}$ is upper-bounded by the maximum value of $g_1$ (the gaussian pdf with identity covariance matrix).

For $\sigma_d \neq 1$ the problem can be reduced to the previous case via the following scaling: Consider $\mathcal{P}(x) = \frac{1}{\sigma_d}\mathcal{D}_{\sigma_d}(\sigma_d x)$. Then $\mathcal{P}(y) = E[\tilde{X}|\tilde{Y} = y]$ is an MMSE denoiser with variance 1 with $\tilde{X} = X/\sigma_d$ and $\tilde{Y} = \tilde{X} + N$. So we can find (according to the previous argument for $\sigma_d = 1$) $f$ such that $\mathcal{P} = \text{prox}_f$. Applying a change of variables in the proximal operator we obtain

$$\mathcal{D}_{\sigma_d}(y) = \sigma_d \mathcal{P}(y/\sigma_d) = \text{prox}_{\phi_{\sigma_d}}(y)$$

where

$$\phi_{\sigma_d}(x) = \sigma_d^2 f(x/\sigma_d)$$

Finally, since $f$ is lower-bounded $\phi_{\sigma_d}$ is lower bounded too.

□

# C Convergence to critical point

Since we are optimizing

$$E(x) = g(Hx) + f(x)$$

we would like to show that

$$\lim_{k\to\infty} \nabla E(x_k) = 0$$

We can almost conclude this from Theorem 1. Indeed

$$\nabla E(x_k) = H^* \nabla g(Hx) + \nabla f(x)$$

From Theorem 1 we have that

$$\nabla_w \mathcal{L}_\beta = z_k - Hx_k \to 0$$

$$\nabla_z \mathcal{L}_\beta = w_k + \beta(z - Hx) + \nabla g(z) \to 0$$

$$\nabla_x \mathcal{L}_\beta = \nabla f(x_k) - H^* w_k + \beta H^*(Hx_k - z_k) \to 0$$

Putting all together we have:

$$\begin{aligned} z_k - Hx_k &\to 0 \\ \nabla g(z_k) + w_k &\to 0 \\ \nabla f(x_k) - H^* w_k &\to 0 \end{aligned}$$

Since $\nabla g$ is continuous we get that

$$\nabla g(Hx_k) + w_k \to \nabla g(z_k) + w_k \to 0$$

This means that

$$\begin{aligned}\nabla E(x_k) &= H^*\nabla g(Hx_k) + \nabla f(x_k) \\ &= H^*\nabla g(Hx_k) + H^*w_k + \nabla f(x_k) - H^*w_k \\ &= H^*(\nabla g(Hx_k) + w_k) + (\nabla f(x_k) - H^*w_k) \\ &\to 0\end{aligned}$$

**Conclusion:** If $f$ is differentiable, then Theorem 1 implies that the Linearized ADMM converges to a critical point of the original objective $E(x)$.