Fractional-gradient Sparsity with Autoencoding Sequential Deep Image Prior for 3D CT Reconstruction

Supplementary Material

7. Appendix: Theoretical Proofs

In this section, we provide detailed proofs for the theoretical claims in the main manuscript.

7.1. Proof of Lemma 1 (Majorization of the Smoothed Numerator)

Proof. We consider the smoothed numerator

$$N_\delta(\mathbf{z}) = \sum_i \sqrt{|(\nabla \mathbf{z})_i|^2 + \delta}, \tag{15}$$

where $\delta > 0$ is a smoothing parameter.

For each spatial index i, define the scalar function

$$h(t) = \sqrt{t + \delta}, \qquad t \geq 0. \tag{16}$$

Its first and second derivatives are

$$h'(t) = \frac{1}{2\sqrt{t+\delta}}, \qquad h''(t) = -\frac{1}{4}(t+\delta)^{-3/2} < 0. \tag{17}$$

Hence, h is strictly concave on $[0, \infty)$.

By the supporting-hyperplane property of concave functions, for any $t, t_0 \geq 0$,

$$h(t) \leq h(t_0) + h'(t_0)(t - t_0). \tag{18}$$

Substituting the expression of h, we obtain

$$\sqrt{t+\delta} \leq \sqrt{t_0 + \delta} + \frac{1}{2\sqrt{t_0+\delta}}(t - t_0). \tag{19}$$

Now let

$$t = |(\nabla \mathbf{z})_i|^2, \qquad t_0 = |(\nabla \mathbf{z}^{k-1})_i|^2. \tag{20}$$

Then, for each i,

$$\sqrt{|(\nabla \mathbf{z})_i|^2 + \delta} \leq \sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta} + \frac{|(\nabla \mathbf{z})_i|^2 - |(\nabla \mathbf{z}^{k-1})_i|^2}{2\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}}. \tag{21}$$

Summing the above inequality over all indices i gives

$$\begin{aligned} N_\delta(\mathbf{z}) &= \sum_i \sqrt{|(\nabla \mathbf{z})_i|^2 + \delta} \\ &\leq \sum_i \left[\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta} + \frac{|(\nabla \mathbf{z})_i|^2 - |(\nabla \mathbf{z}^{k-1})_i|^2}{2\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} \right] \\ &= \frac{1}{2}\sum_i \frac{|(\nabla \mathbf{z})_i|^2}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} + C^{k-1}, \end{aligned} \tag{22}$$

where

$$C^{k-1} = \sum_i \left(\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta} - \frac{|(\nabla \mathbf{z}^{k-1})_i|^2}{2\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} \right) \tag{23}$$

is independent of $\mathbf{z}$. Therefore, defining

$$\tilde{N}_\delta(\mathbf{z} \mid \mathbf{z}^{k-1}) = \frac{1}{2}\sum_i \frac{|(\nabla \mathbf{z})_i|^2}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} + C^{k-1}, \tag{24}$$

we have

$$N_\delta(\mathbf{z}) \leq \tilde{N}_\delta(\mathbf{z} \mid \mathbf{z}^{k-1}), \tag{25}$$

which proves the upper-bound property.

Next, we verify the tangency condition. Setting $\mathbf{z} = \mathbf{z}^{k-1}$, we have

$$\begin{aligned} \tilde{N}_\delta(\mathbf{z}^{k-1} \mid \mathbf{z}^{k-1}) &= \frac{1}{2}\sum_i \frac{|(\nabla \mathbf{z}^{k-1})_i|^2}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} + C^{k-1} \\ &= \sum_i \sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta} = N_\delta(\mathbf{z}^{k-1}). \end{aligned} \tag{26}$$

Hence the tangency condition also holds.

Finally, by defining the diagonal weight matrix

$$\mathbf{W}^{k-1}_{i,i} = \frac{1}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}}, \tag{27}$$

the quadratic part of the surrogate can be written compactly as

$$\frac{1}{2}\sum_i \frac{|(\nabla \mathbf{z})_i|^2}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} = \frac{1}{2}(\nabla \mathbf{z})^\top \mathbf{W}^{k-1}(\nabla \mathbf{z}). \tag{28}$$

This completes the proof. □

7.2. Proof of Lemma 2 (Lipschitz Continuous Gradient)

Proof. For fixed ϕ^k and $\mathbf{z}^{k-1}$, the surrogate objective is

$$\mathcal{L}_s(\phi^k, \mathbf{z} \mid \mathbf{z}^{k-1}) = F(\phi^k, \mathbf{z}) + S(\mathbf{z} \mid \mathbf{z}^{k-1}), \tag{29}$$

where

$$S(\mathbf{z} \mid \mathbf{z}^{k-1}) = \frac{\gamma}{M^{k-1}} \tilde{N}_\delta(\mathbf{z} \mid \mathbf{z}^{k-1}), \;\; M^{k-1} = \|\nabla \mathbf{z}^{k-1}\|_2 + \epsilon. \tag{30}$$

We show that each term has Lipschitz continuous gradient with respect to $\mathbf{z}$ on the bounded iterate set.

Distribution Statement A. Approved for public release: distribution is unlimited.

First, since $S(\mathbf{z} \mid \mathbf{z}^{k-1})$ is quadratic in $\mathbf{z}$, its Hessian is constant and given by

$$\nabla^2_{\mathbf{z}} S(\mathbf{z} \mid \mathbf{z}^{k-1}) = \frac{\gamma}{M^{k-1}} \nabla^\top \mathbf{W}^{k-1} \nabla. \tag{31}$$

Because $\epsilon > 0$, we have $M^{k-1} \geq \epsilon$. Moreover, since $\delta > 0$,

$$\mathbf{W}^{k-1}_{i,i} = \frac{1}{\sqrt{|(\nabla \mathbf{z}^{k-1})_i|^2 + \delta}} \leq \frac{1}{\sqrt{\delta}}. \tag{32}$$

Hence

$$\left\| \nabla^2_{\mathbf{z}} S(\mathbf{z} \mid \mathbf{z}^{k-1}) \right\|_2 \leq \frac{\gamma}{\epsilon \sqrt{\delta}} \|\nabla^\top \nabla\|_2, \tag{33}$$

which is finite because the discrete gradient operator is linear and bounded in finite dimensions.

Next, consider

$$F(\phi^k, \mathbf{z}) = \|\mathbf{A} f_{\phi^k}(\mathbf{z}) - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z} - f_{\phi^k}(\mathbf{z})\|_2^2. \tag{34}$$

By assumption, for fixed ϕ^k, the mapping $f_{\phi^k}(\mathbf{z})$ is continuously differentiable and has bounded Jacobian and Hessian on the bounded iterate set. Since $\mathbf{A}$ is linear and bounded, standard chain-rule calculations imply that the Hessian $\nabla^2_{\mathbf{z}} F(\phi^k, \mathbf{z})$ is bounded on the same set. Therefore, there exists a constant $L_F > 0$ such that

$$\|\nabla^2_{\mathbf{z}} F(\phi^k, \mathbf{z})\|_2 \leq L_F. \tag{35}$$

Combining the two bounds, we obtain

$$\|\nabla^2_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z} \mid \mathbf{z}^{k-1})\|_2 \leq L_F + \frac{\gamma}{\epsilon \sqrt{\delta}} \|\nabla^\top \nabla\|_2 =: L. \tag{36}$$

Hence the gradient $\nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z} \mid \mathbf{z}^{k-1})$ is L-Lipschitz continuous on the bounded iterate set. □

7.3. Proof of Proposition 1 (Monotonic Descent of the Surrogate Objective)

Proof. By Lemma 2, the gradient of the surrogate objective

$$\mathbf{z} \mapsto \mathcal{L}_s(\phi^k, \mathbf{z} \mid \mathbf{z}^{k-1})$$

is L-Lipschitz continuous. Therefore, by the standard descent lemma, for any $\mathbf{u}, \mathbf{v}$,

$$\begin{aligned}\mathcal{L}_s(\phi^k, \mathbf{u} \mid \mathbf{z}^{k-1}) \leq& \mathcal{L}_s(\phi^k, \mathbf{v} \mid \mathbf{z}^{k-1}) + \\ & \langle \nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{v} \mid \mathbf{z}^{k-1}), \mathbf{u} - \mathbf{v} \rangle + \frac{L}{2} \|\mathbf{u} - \mathbf{v}\|_2^2. \end{aligned} \tag{37}$$

Now set $\mathbf{v} = \mathbf{z}^{k-1}, \mathbf{u} = \mathbf{z}^k = \mathbf{z}^{k-1} - \beta \nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z} \mid \mathbf{z}^{k-1})\big|_{\mathbf{z}=\mathbf{z}^{k-1}}$. Then

$$\mathbf{z}^k - \mathbf{z}^{k-1} = -\beta \nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1}). \tag{38}$$

Substituting into the descent inequality yields

$$\begin{aligned} & \mathcal{L}_s(\phi^k, \mathbf{z}^k \mid \mathbf{z}^{k-1}) \\ \leq & \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1}) - \beta \|\nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1})\|_2^2 \\ & + \frac{L\beta^2}{2} \|\nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1})\|_2^2 \\ = & \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1}) - \left(\beta - \frac{L\beta^2}{2}\right) \|\nabla_{\mathbf{z}} \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1})\|_2^2. \end{aligned} \tag{39}$$

If $\beta \leq 1/L$, then

$$\beta - \frac{L\beta^2}{2} \geq \frac{\beta}{2} > 0, \tag{40}$$

and therefore

$$\mathcal{L}_s(\phi^k, \mathbf{z}^k \mid \mathbf{z}^{k-1}) \leq \mathcal{L}_s(\phi^k, \mathbf{z}^{k-1} \mid \mathbf{z}^{k-1}). \tag{41}$$

This proves the monotonic descent of the surrogate objective. □

7.4. Proof Sketch of Theorem 1 (Subsequence Convergence under KL Framework)

Proof. The proof follows a standard KL-based argument for surrogate-based alternating optimization.

First, by Proposition 1, the $\mathbf{z}$-update yields a sufficient decrease of the surrogate objective at each iteration. By assumption, the ϕ-update together with the $\mathbf{z}$-update defines a bounded surrogate-based alternating sequence that satisfies a sufficient decrease condition and a relative error condition.

Second, the involved functions are semi-algebraic. Indeed, the data fidelity term is polynomial in its arguments up to composition with the network mapping; the autoencoding term is quadratic; and the surrogate fractional term is a weighted quadratic function with weights determined from the previous iterate. Under the stated assumptions, the resulting surrogate-based objective belongs to the class of semi-algebraic (more generally, KL) functions.

Therefore, the abstract convergence results for KL functions apply: the sufficient decrease condition prevents oscillatory growth of the surrogate objective, the relative error condition controls the first-order residual by successive differences, and boundedness ensures the existence of accumulation points. Consequently, the sequence has finite length,

$$\sum_{k=1}^{\infty} \|(\phi^{k+1}, \mathbf{z}^{k+1}) - (\phi^k, \mathbf{z}^k)\|_2 < \infty, \tag{42}$$

and every accumulation point is a critical point of the surrogate-based alternating scheme.

We emphasize that this theorem concerns the surrogate-based alternating scheme induced by the IRLS-type approximation, rather than a strict global majorization of the original full ratio objective. □