A Bilevel Sensitivity-Corrected Reconstruction Framework with Deep Priors for Parallel MRI

Supplementary Material

I. Convergence Proof for SRSC

In this section, we provide a rigorous convergence proof for the proposed SRSC model. We analyze the iterative process used to solve the bilevel optimization problem and establish the conditions under which the algorithm converges. Specifically, we show that the alternating scheme employed in model (6) converges to a stationary point under appropriate assumptions.

We define the following joint energy (or surrogate objective) in (u, x):

$$\begin{aligned}\mathcal{J}(u,x) :=&\frac{1}{2}\left\|\mathcal{P}\mathcal{F}\mathcal{S}(x)u-g\right\|_2^2+\left\|\Gamma W\mathcal{F}^\top\left(\mathcal{Q}\mathcal{F}\mathcal{S}(x)u+g\right)\right\|_1+\\&\frac{1}{2}\left\|(G-I)(\mathcal{Q}\mathcal{F}x+g)\right\|_2^2+\frac{\eta}{2}\left\|x-\mathcal{S}(x)u\right\|_2^2.\end{aligned}\tag{1}$$

For fixed x, the first two terms in (1) reduce to a SENSE3d-type objective in u. For fixed u, the last two terms form a SPIRiT-based calibration problem in x with a quadratic coupling to $\mathcal{S}(x)u$. In this sense, $\mathcal{J}$ is consistent with the bilevel formulation discussed in model (6).

We collect the assumptions used in the analysis.

Assumption A1 (bounded linear operators). All linear operators $\mathcal{P}, \mathcal{Q}, \mathcal{F}, W, G$ are bounded, and $\mathcal{F}$ is unitary (up to normalization). The SPIRiT operator G is such that $G-I$ is bounded.

Assumption A2 (regularity of the CSMs mapping). The CSM mapping $\mathcal{S}(\cdot)$ is continuously differentiable and locally Lipschitz on the region of interest; in particular, whenever $\sum_{\ell=1}^{c}|x_\ell[i]|^2$ is bounded away from zero, the entries of $\mathcal{S}(x)$ and their first derivatives are bounded.

Assumption A3 (coercivity and KL property). The function $\mathcal{J}$ is coercive, i.e.,

$$\|(u,x)\|_2\to\infty \quad\Longrightarrow\quad \mathcal{J}(u,x)\to+\infty,$$

and $\mathcal{J}$ satisfies the Kurdyka–Lojasiewicz (KL) property at every point of its domain.

Assumption A3 is standard in KL-based convergence analysis. In our setting, $\mathcal{J}$ is a finite sum of compositions of polynomials, rational, and ℓ_1-type functions with linear and smooth mappings. This makes $\mathcal{J}$ semi-algebraic (or more generally definable in an o-minimal structure), hence it has the KL property; see, e.g., [1].

We now record some basic consequences of these assumptions.

Lemma 1 (Basic properties of $\mathcal{J}$)**.** *Under Assumptions A1–A3, the function $\mathcal{J}:\mathbb{C}^n\times\mathbb{C}^{cn}\to\mathbb{R}\cup\{+\infty\}$ defined in* (1) *satisfies:*

1) *$\mathcal{J}$ is proper and lower semicontinuous;*
2) *Every sublevel set $\{(u,x):\mathcal{J}(u,x)\le a\}$ is bounded for any $a\in\mathbb{R}$;*
3) *$\mathcal{J}$ has the KL property at every point of its domain.*

Proof. Properness and lower semicontinuity follow from continuity of the quadratic terms and lower semicontinuity of the ℓ_1-norm, plus Assumption A2 on $\mathcal{S}(\cdot)$. Coercivity in Assumption A3 implies boundedness of sublevel sets. The KL property follows from the semi-algebraic (or definable) structure of $\mathcal{J}$; see, for example, [1]. □

For the convergence analysis, we consider an *idealized* SRSC scheme in which each subproblem is solved exactly. Starting from an initial pair (u^0, x^0), the algorithm generates a sequence $\{(u^k, x^k)\}_{k\in\mathbb{N}}$ as follows:

Algorithm A.1 (Idealized SRSC). For $k = 0, 1, 2, \dots$

1) **(u-step)** Given x^k, set
$$u^{k+1} \in \arg\min_{u\in\mathbb{C}^n} \mathcal{J}(u, x^k).$$
This is a convex but nonsmooth problem in u (SENSE3d-type objective).
2) **(x-step)** Given u^{k+1}, set
$$x^{k+1} \in \arg\min_{x\in\mathbb{C}^{cn}} \mathcal{J}(u^{k+1}, x).$$
This is a smooth, strongly convex problem in x (SPIRiT calibration plus quadratic coupling) with a unique minimizer.

In the implemented Algorithm 1, the u-step is carried out by PD3O iterations and the x-step by a few conjugate gradient iterations on the linear system associated with (1). For the theoretical analysis, we work with Algorithm A.1 and then briefly comment on the practical scheme at the end of this appendix.

Lemma 2 (Well-posedness of the block subproblems)**.** *Under Assumptions A1–A3, the following hold for every fixed k:*

1) *For any fixed x^k, the u-subproblem $u \mapsto \mathcal{J}(u, x^k)$ is convex and has at least one minimizer. If in addition the operator $\mathcal{PFS}(x^k)$ has trivial null space (up to constants), the minimizer is unique.*
2) *For any fixed u^{k+1}, the x-subproblem $x \mapsto \mathcal{J}(u^{k+1}, x)$ is smooth and strongly convex, hence has a unique minimizer.*

Proof. For fixed x^k, the terms in $\mathcal{J}(u, x^k)$ involving u consist of a quadratic data-fidelity term and an ℓ_1-norm composed with a linear operator, which is convex. Coercivity implies the existence of minimizers, and uniqueness follows from the strict convexity of the quadratic part under mild conditions on $\mathcal{PFS}(x^k)$.

For fixed u^{k+1}, the terms in $\mathcal{J}(u^{k+1}, x)$ involving x are smooth quadratics: $\frac{1}{2}\|(G-I)(\mathcal{QF}x+g)\|_2^2$ and $\frac{\eta}{2}\|x-\mathcal{S}(x^k)u^{k+1}\|_2^2$. The first term defines a positive semidefinite quadratic form and the second adds strict convexity because $\eta > 0$. Consequently, $x \mapsto \mathcal{J}(u^{k+1}, x)$ is strongly convex, whence it has a unique minimizer. □

The next lemma records the basic descent property of Algorithm A.1.

Lemma 3 (Descent and finite-length property)**.** *Let $\{(u^k, x^k)\}$ be the sequence generated by Algorithm A.1. Under Assumptions A1–A3, we have:*

1) *$\mathcal{J}(u^{k+1}, x^{k+1}) \le \mathcal{J}(u^k, x^k)$ for all k, and the sequence $\{\mathcal{J}(u^k, x^k)\}$ is convergent;*
2) *The sequence $\{(u^k, x^k)\}$ has bounded sublevel sets and satisfies*
$$\sum_{k=0}^{\infty}\Big(\|u^{k+1} - u^k\|_2^2 + \|x^{k+1} - x^k\|_2^2\Big) < +\infty.$$

Proof. By definition of u^{k+1},
$$\mathcal{J}(u^{k+1}, x^k) \le \mathcal{J}(u^k, x^k),$$

since u^{k+1} minimizes $\mathcal{J}(\cdot, x^k)$. Similarly, x^{k+1} minimizes $\mathcal{J}(u^{k+1}, \cdot)$, so

$$\mathcal{J}(u^{k+1}, x^{k+1}) \le \mathcal{J}(u^{k+1}, x^k).$$

Combining the two inequalities yields

$$\mathcal{J}(u^{k+1}, x^{k+1}) \le \mathcal{J}(u^k, x^k), \qquad \forall k.$$

By Lemma 1 and Assumption A3, $\mathcal{J}$ is bounded from below, so $\{\mathcal{J}(u^k, x^k)\}$ converges to some limit $\mathcal{J}^\star$.

The finite-length property follows from standard arguments in the analysis of block coordinate descent for KL functions; see, e.g., [2, Theorem 3.2] and [3, Theorem 3.1]. Intuitively, the strict convexity of each subproblem yields a sufficient decrease condition $\mathcal{J}(u^k, x^k) - \mathcal{J}(u^{k+1}, x^{k+1}) \ge c\|(u^{k+1}, x^{k+1}) - (u^k, x^k)\|^2$ for some $c > 0$, from which the summability of the squared increments follows. □

We are now ready to state the main convergence result for the idealized SRSC algorithm.

Theorem 1 (Convergence of Algorithm A.1)**.** *Let Assumptions A1–A3 hold, and let* $\{(u^k, x^k)\}$ *be the sequence generated by Algorithm A.1. Then:*

1) *The sequence* $\{(u^k, x^k)\}$ *is bounded and has finite length, i.e.,* $\sum_{k=0}^{\infty} \|(u^{k+1}, x^{k+1}) - (u^k, x^k)\|_2 < +\infty$.
2) *The sequence* $\{(u^k, x^k)\}$ *converges to a limit* $(u^\star, x^\star)$.
3) *The limit* $(u^\star, x^\star)$ *is a critical point of* $\mathcal{J}$*, in the sense that*

$$0 \in \partial \mathcal{J}(u^\star, x^\star),$$

where ∂ *denotes the limiting subdifferential. Equivalently,* $(u^\star, x^\star)$ *satisfies the first-order optimality conditions of the SENSE3d subproblem in* u *and the SPIRiT-calibration subproblem in* x.

Proof. Boundedness of the sequence and the finite-length property follow from Lemma 1 and Lemma 3. The KL property of $\mathcal{J}$ and the sufficient decrease and relative error conditions implied by the exact minimization of each block (Lemma 2) place Algorithm A.1 within the framework of block coordinate descent on KL functions. The conclusion then follows from standard results on the global convergence of such methods, see, for example, [2, Theorem 3.2] and [3, Theorem 3.1]. □

The implemented SRSC algorithm differs from Algorithm A.1 in two respects:

- the u-subproblem is solved approximately by a fixed number of PD3O iterations with fixed $\mathcal{S}$ and Γ;
- the x-subproblem is solved approximately by a fixed number of conjugate gradient iterations on the normal equations.

In practice, these inexact solves still produce a decreasing sequence of energies and empirically stable behavior. A fully rigorous treatment of the inexact case can be obtained by invoking KL-based convergence results for inexact block coordinate or PALM-type methods, provided that the inner errors are suitably controlled (e.g., summable). For clarity of presentation, we restrict the formal analysis here to the idealized exact-solve setting.

II. SRSC WITH RED

When the PnP-based Regularization by Denoising (RED) is adopted and the explicit classical linear regularizer is replaced by a deep denoiser, the bilevel SRSC model (6) degenerates into:

$$\begin{cases} \tilde{u} \in \arg\min\limits_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathcal{P}\mathcal{F}\mathcal{S}(\tilde{x})u - g\|_2^2 + \lambda \mathcal{R}(u) \right\}, \\ \hat{x} = \mathcal{S}(\tilde{x})\, \tilde{u}, \\ \tilde{x} \in \arg\min\limits_{x \in \mathbb{C}^{cn}} \left\{ \frac{1}{2} \|(G - I)(\mathcal{Q}\mathcal{F}x + g)\|_2^2 + \frac{\eta}{2} \|x - \hat{x}\|_2^2 \right\}. \end{cases} \tag{2}$$

Here, λ denotes the regularization parameter and $\mathcal{R}(\cdot)$ represents an implicitly defined regularization function. The upper-level u-subproblem can be solved using the ADMM-PnP method [4]–[6]. By constructing the augmented Lagrangian and performing variable splitting, the iterative update process can be formulated as follows:

$$v^{k+1} = \arg\min_{v} \left\{ \lambda \mathcal{R}(v) + \frac{\rho_k}{2} \|v - \hat{v}^k\|_2^2 \right\}, \tag{3a}$$

$$u^{k+1} = \arg\min_{u} \left\{ \frac{1}{2} \|\mathcal{P}\mathcal{F}\mathcal{S}(\tilde{x})u - g\|_2^2 + \frac{\rho_k}{2} \|u - \hat{u}^k\|_2^2 \right\}, \tag{3b}$$

$$\alpha^{k+1} = \alpha^k + u^{k+1} - v^{k+1}. \tag{3c}$$

Here, $\hat{v}^k = u^k + \alpha^k$ and $\hat{u}^k = v^{k+1} - \alpha^k$. Based on the RED framework, the v-subproblem (3a) can be solved using a deep denoiser. To ensure fairness in the ablation study, we employ the same SGM-based denoiser for reconstruction, i.e.,

$$v^{k+1} = \mathcal{N}_\Theta(\hat{v}^k).$$

For the u-subproblem (3b), its smooth and differentiable form allows for an efficient solution using the Conjugate Gradient (CG) method [7], which corresponds to solving the following linear system:

$$(\mathcal{S}^\top \mathcal{F}^\top \mathcal{P}\mathcal{F}\mathcal{S} + \rho_k I)u = \rho_k \hat{u}^k + \mathcal{S}^\top \mathcal{F}^\top g. \tag{4}$$

Based on the above preparation, an approximate solution to the upper-level subproblem can be obtained. The SRSC model with RED regularization is summarized in Algorithm 1, and we set $\rho_k \equiv 1$ to eliminate the need for tuning this hyperparameter.

Algorithm 1 The SRSC with RED algorithm for solving model (2)

```
Ensure: v̂^k = u^k + α^k and û^k = v^{k+1} − α^k.
for k = 0, 1, ... do
  // SPIRiT-based Calibration for CSMs:
  if k mod 10 = 0 and k > 0 then
    x̂ = S u^k
    x̃ = CG(x̂)
    S = S(x̃)
  end if
  v^{k+1} = N_Θ(v̂^k)
  u^{k+1} = Use CG to solve linear system (4)
  α^{k+1} = α^k + u^{k+1} − v^{k+1}
end for
```

REFERENCES

[1] H. Attouch, J. Bolte, and P. Redont, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, pp. 438–457, 2010.

[2] H. Attouch and J. Bolte, "On the convergence of the proximal algorithm for nonsmooth functions involving analytic features," *Mathematical Programming, Ser. A*, vol. 116, pp. 5–16, 2009.

[3] J. Bolte, S. Sabach, and M. Teboulle, “Proximal alternating linearized minimization for nonconvex and nonsmooth problems,” *Mathematical Programming, Ser. A*, vol. 146, pp. 459–494, 2014.

[4] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[5] S. H. Chan, X. Wang, and O. A. Elgendy, “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016.

[6] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, “Denoising prior driven deep neural network for image restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2305–2318, 2018.

[7] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 2006.