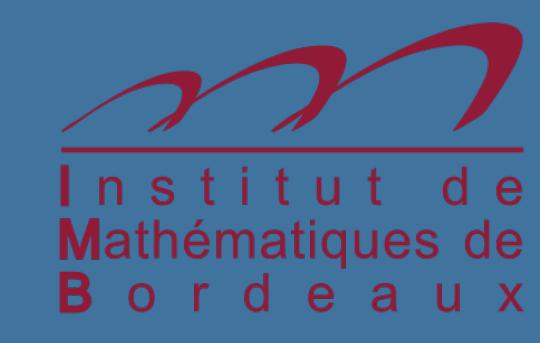
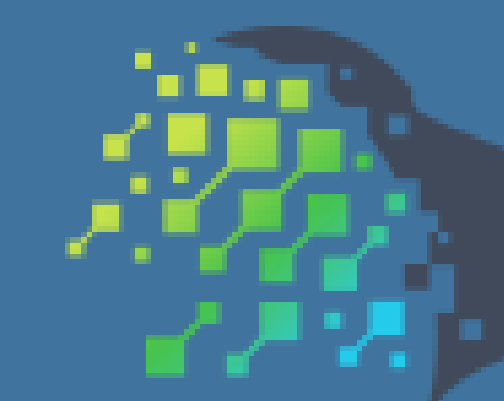


Provably Accelerated Imaging with Restarted Inertia and Score-based Image Priors



M. Renaud¹ J. Hermant¹ D. Wei² Y. Sun²

¹IMB, Univ. Bordeaux ²Johns Hopkins University



ICLR

Background

Inverse problem.

$$y = A(x) + n,$$

where $x \in \mathbb{R}^d$ is the unknown signal, $y \in \mathbb{R}^m$ the observation, $A : \mathbb{R}^d \rightarrow \mathbb{R}^m$ the forward operator, and n measurement noise.

MAP estimation. A common approach to solve inverse problems is *maximum a posteriori* (MAP) estimation:

$$\hat{x} \in \arg \max_x p(y|x)p(x) = \arg \min_{x \in \mathbb{R}^d} \{F(x) := f(x) + g(x)\}$$

- $f(x) = -\log p(y|x)$: data fidelity
- $g(x) = -\log p(x)$: prior (regularization)

F can be minimized with optimization algorithms.

Regularization by Denoising (RED). [2]

Approximate the prior gradient using the score $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$:

$$\nabla g(x) \approx -S(x), \quad S(x) := -(x - D_\sigma(x))/\sigma^2$$

for a denoiser $D_\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with noise strength σ^2 .

Gradient step

$$x^+ = x - \eta(\nabla f(x) - S(x)) \quad (\text{RED-GM})$$

Proximal gradient

$$x^+ = \text{prox}_{\eta f}(x + \eta S(x)) \quad (\text{RED-Prox})$$

where $\text{prox}_{\eta f}(z) := \arg \min_{x \in \mathbb{R}^d} \{\frac{1}{2}\|x - z\|^2 + \eta f(x)\}$.

Drawback: For each measurement y , need to run an iterative algorithm. **Accelerating the convergence ?**

RISP for Accelerated Imaging

Inertia + Restart has proven to be efficient for highly non-convex problems [1]. Introducing *Restarted Inertia with Score-based Priors* (RISP), we augment score-based algorithms with:

- **Inertia** as an **accelerating** mechanism.
- **Restart** as a **stabilizing** mechanism.

Algorithm 1 RISP

Require: $x^{-1} = x^0 \in \mathbb{R}^d$, $K \in \mathbb{N}$, $\eta > 0$, $\theta \in (0, 1]$, and $B \geq 0$

- 1: **while** $0 \leq k < K$ **do**
- 2: $z^k = x^k + (1 - \theta)(x^k - x^{k-1})$ ▷ inertia
- 3: Option RISP-GM: $x^{k+1} = z^k - \eta(\nabla f(z^k) - S(z^k))$
- 4: Option RISP-Prox: $x^{k+1} = \text{prox}_{\eta f}(z^k + \eta S(z^k))$
- 5: $k = k + 1$
- 6: **if** $k \sum_{t=0}^{k-1} \|x^{t+1} - x^t\|^2 > B^2$ **then** ▷ restart
- 7: $x^{-1} = x^0 = x^k$, $k = 0$
- 8: **end if**
- 9: **end while**
- 10: $K_0 = \arg \min_{\lfloor \frac{K}{2} \rfloor \leq k \leq K-1} \|x^{k+1} - x^k\|$
- 11: **return** $\hat{z} = \frac{1}{K_0+1} \sum_{k=0}^{K_0} z^k$

Convergence Analysis of RISP

Assumption 1. The score function $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a gradient field, *i.e.*, there exists a regularizer g such that $S = -\nabla g$.

Assumption 2. f has a Lipschitz continuous gradient, S is Lipschitz continuous.

Assumption 3. f has Lipschitz continuous Hessian, S is Lipschitz Jacobian continuous.

Assumption 4. f is convex, g is ν -weakly convex, *i.e.*, $g + \frac{\nu}{2}\|\cdot\|^2$ is convex.

Proposition. (Convergence of RED) Under Assumptions 1 and 2, (RED-GM) or (RED-Prox) with n iterations outputs a point \hat{x} such that

$$\|\nabla F(\hat{x})\| = \mathcal{O}(n^{-1/2}),$$

Theorem. Under Assumptions 1, 2 and 3, RISP (Option RISP-GM) with n iterations outputs a point \hat{x} such that

$$\|\nabla F(\hat{x})\| = \mathcal{O}(n^{-4/7}).$$

Same result for Option RISP-Prox if adding Assumption 4.

References

- [1] Huan Li and Zhouchen Lin. Restarted nonconvex accelerated gradient descent: No more poly-logarithmic factor in the $\mathcal{O}(\epsilon^{-7/4})$ complexity. *Journal of Machine Learning Research*, 24(157):1–37, 2023.
- [2] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Numerical experiments

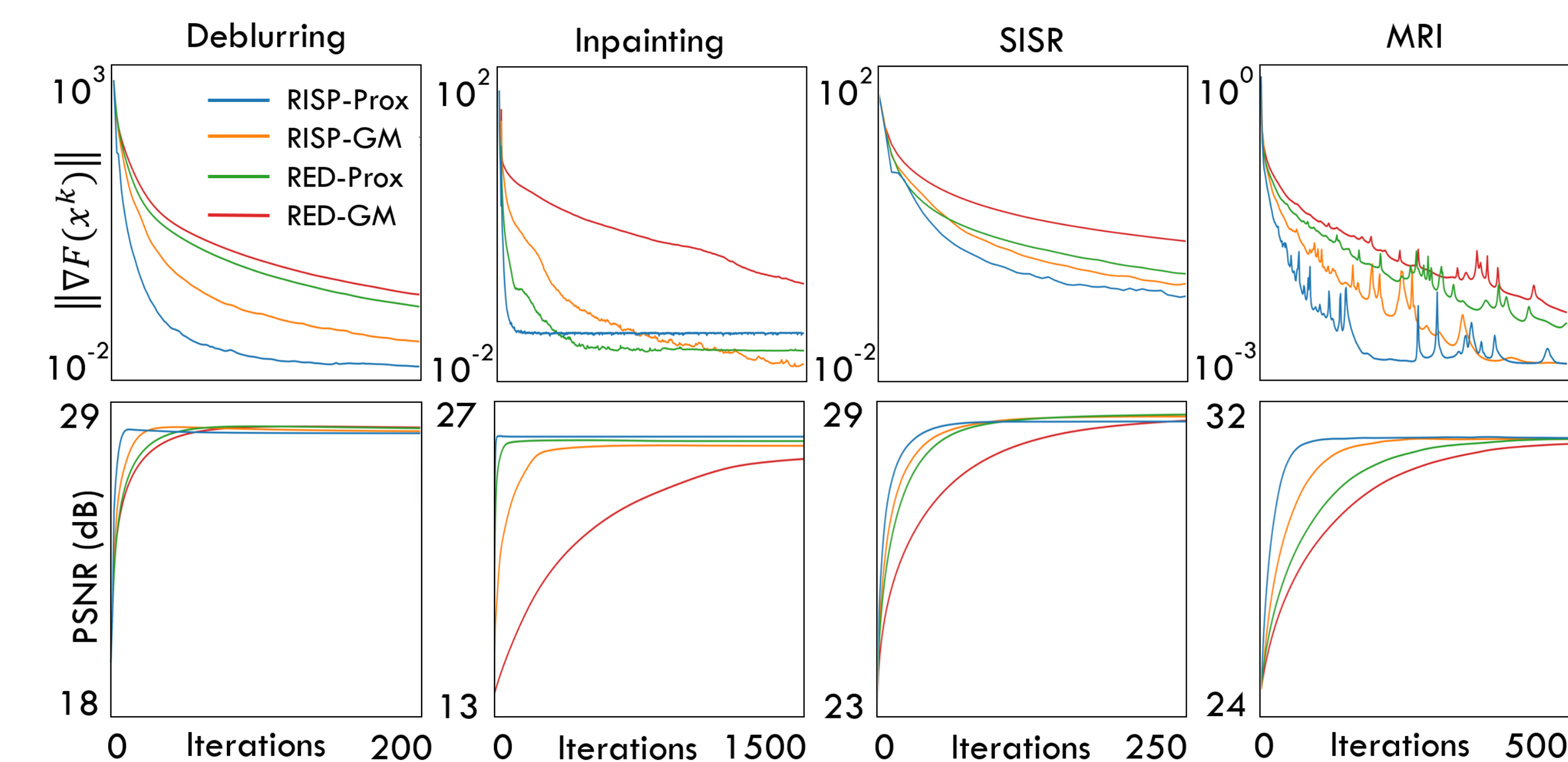


Figure 1. Linear Experiments

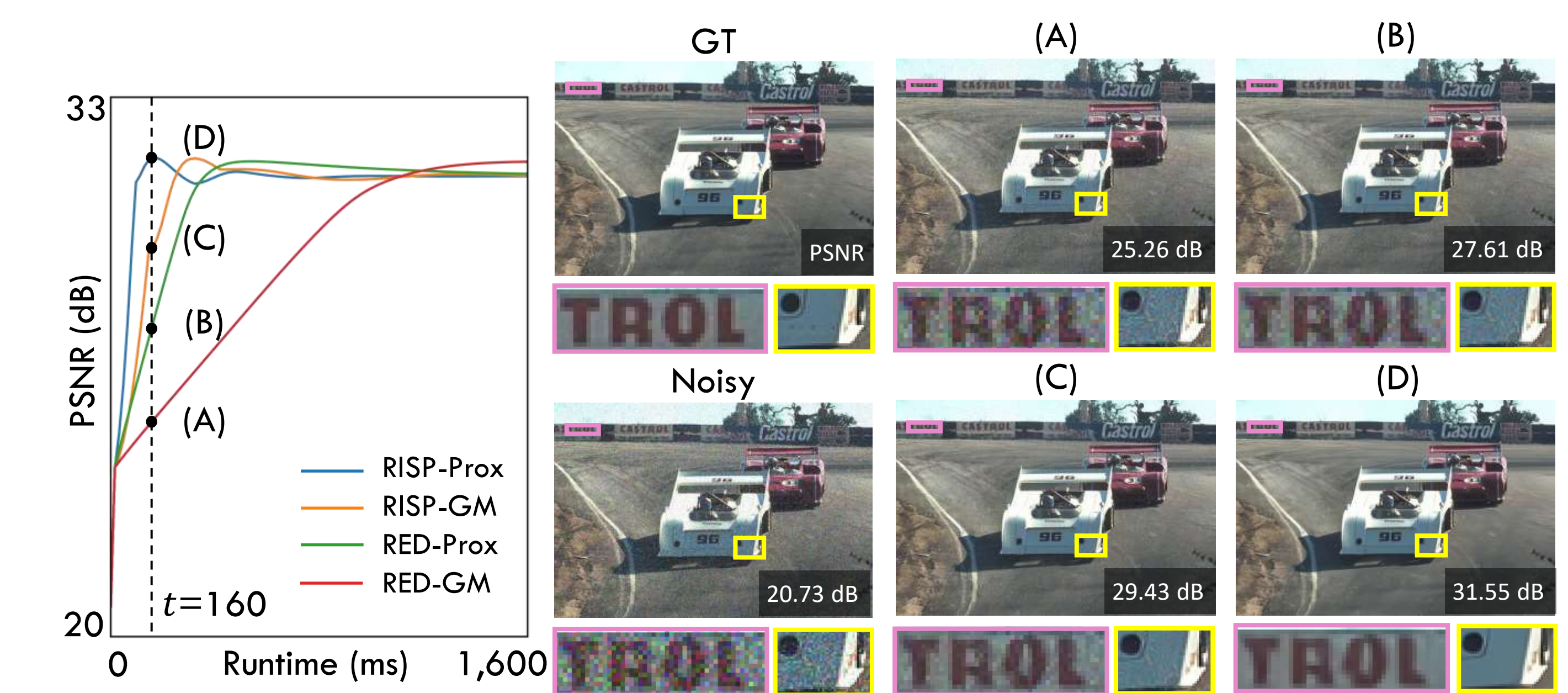


Figure 2. Non-linear experiment: Rician noise removal

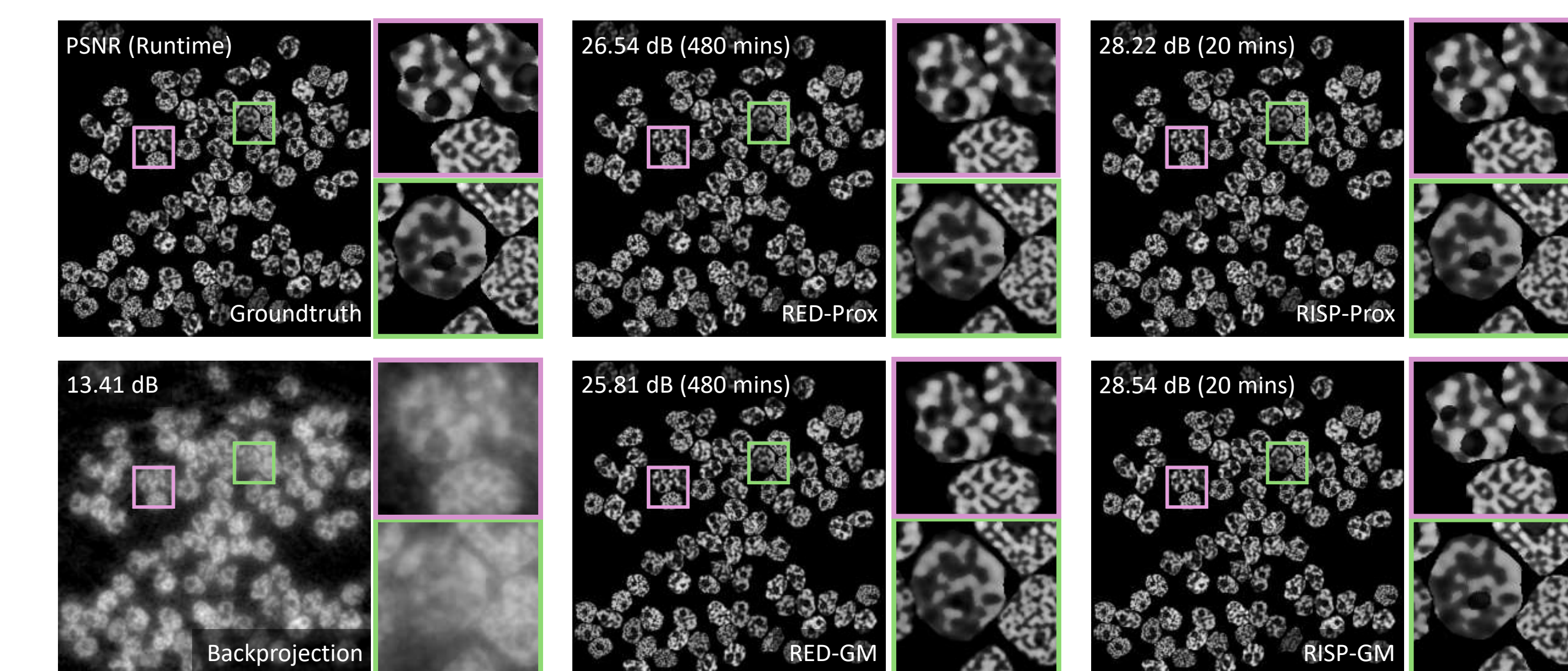


Figure 3. Large scale experiment: inverse scattering