Fast and Provable ADMM for learning with Generative Priors

Fabian Latorre
https://lions.epfl.ch

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland

NeurIPS 2019

[Vancouver, Canada. December 2019]

Joint work with

Armin Eftekhari and Volkan Cevher @ LIONS

lions@epfl HASLERSTIFTUNG Google AI SDSC ZEISS FNSNF Fonds National Suisse Schweizerischer Nationalfonds Fondo Nazionale Svizzero Swiss National Science Foundation Office of Naval Research Global Science & Technology erc EPFL

Classical Signal Recovery

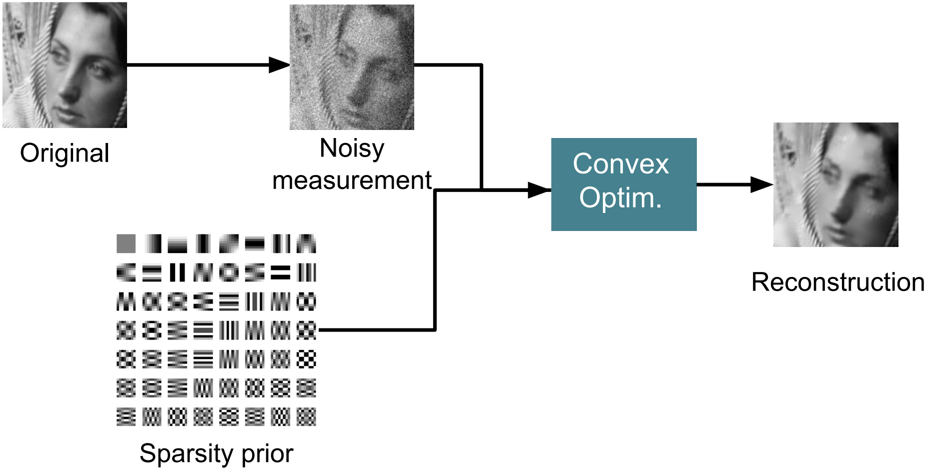

Figure: Recovering a signal with convex optimization and a sparsity prior

Leveraging GANs for Signal Recovery

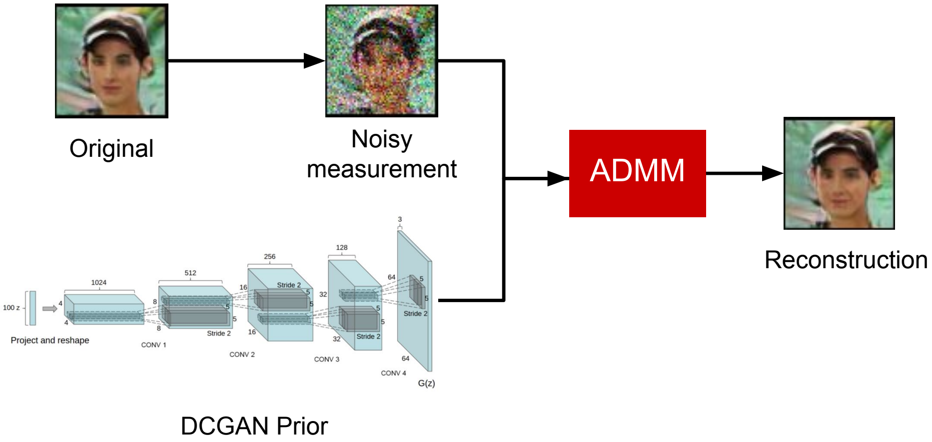

Figure: Recovering a signal with nonconvex optimization and a generative prior.

Optimization Template

$$\min_{w,z} L(w) + R(w) + H(z) \qquad \text{subject to } w = G(z) \tag{1}$$

Optimization Template

$$\min_{w,z} L(w) + R(w) + H(z) \qquad \text{subject to } w = G(z) \tag{1}$$

- L is convex and smooth

Optimization Template

$$\min_{w,z} L(w) + R(w) + H(z) \qquad \text{subject to } w = G(z) \tag{1}$$

- L is convex and smooth
- R, H convex, possibly non-smooth but proximal friendly.

Optimization Template

$$\min_{w,z} L(w) + R(w) + H(z) \qquad \text{subject to } w = G(z) \tag{1}$$

- L is convex and smooth
- R, H convex, possibly non-smooth but proximal friendly.
- G differentiable generative model

Decoupling via alternating minimization / penalty methods

Decoupling via alternating minimization / penalty methods

Definition (Augmented Lagrangian)

Let $\rho > 0$

$$\mathcal{L}_\rho(w, z, \lambda) := L(w) + \langle w - G(z), \lambda \rangle + \frac{\rho}{2}\|w - G(z)\|^2 \tag{2}$$

Decoupling via alternating minimization / penalty methods

Definition (Augmented Lagrangian)

Let $\rho > 0$

$$\mathcal{L}_\rho(w, z, \lambda) := L(w) + \langle w - G(z), \lambda \rangle + \frac{\rho}{2}\|w - G(z)\|^2 \quad (2)$$

Our problem (1) is equivalent to

$$\min_{w,z} \max_{\lambda} \mathcal{L}_\rho(w, z, \lambda) + R(w) + H(z) \quad (3)$$

Iterates of Linearized ADMM

$$
\begin{aligned}
z_{t+1} &\leftarrow P_{\beta H}(z_t - \beta \nabla_z \mathcal{L}_\rho(w_t, z_t, \lambda_t)) \\
w_{t+1} &\leftarrow P_{\alpha R}(w_t - \alpha \nabla_w \mathcal{L}_\rho(w_t, z_{t+1}, \lambda_t)) \\
\lambda_{t+1} &\leftarrow \lambda_t + \sigma_{t+1} \cdot (w_{t+1} - G(z_{t+1}))
\end{aligned}
$$

P_A is the proximal mapping of A.

Example: nonsmooth projections

ℓ_∞ projection

$$\min_{w,z} \|w - w^\natural\|_\infty \quad \text{subject to } w = G(z) \tag{4}$$

$L(w) = H(z) = 0,\ R(w) = \|w - w^\natural\|_\infty$. Proximal mapping is efficient.

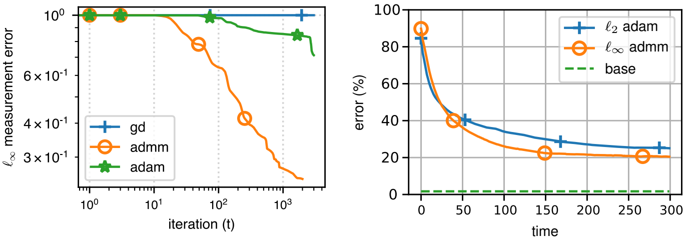

Figure: measurement error per iteration (left). Accuracy on denoised samples (right). MNIST.

Thank you!

5:30 - 07:30 PM @ East Exhibition Hall B + C
Poster #76