Sharp exact penalty formulations in signal recovery

Lijun Ding, University of Wisconsin–Madison
Alex L. Wang, Purdue University

INFORMS Annual Meeting 2023
October 2023

Outline

- Motivation: Sparse recovery
  - $\longrightarrow$ Abstract signal recovery problem (covariance estimation, phase retrieval)
- A new formulation of the abstract problem that is sharp
- Better robustness guarantees
- Faster algorithms

Motivation: Sparse recovery

Sparse recovery setup

- **Recovery task**: Recover $x^\sharp \in \mathbb{R}^n$ from $A \in \mathbb{R}^{m\times n}, b = Ax^\sharp$
- Suppose A entrywise i.i.d. $N(0, 1/m^2)$

$$\left|\mathrm{supp}(x^\sharp)\right| \leq k \ll n \qquad m \asymp k \log(n)$$

- **Convex optimization approach**: In this regime, $x^\sharp$ is unique minimizer of

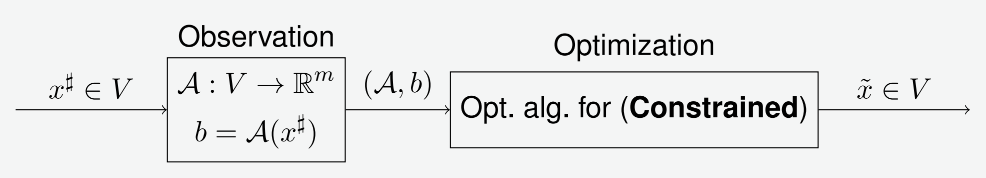
$$\min_{x\in\mathbb{R}^n} \left\{ \|x\|_1 \ : \ Ax = b \right\}$$

Related: Candes and Tao [2005], Recht et al. [2010], Candès et al. [2013]

Abstract signal recovery problem and questions

- $$\min_{x\in\mathbb{R}^n}\left\{\|x\|_1 : \begin{matrix} Ax = b \\ x\in\mathbb{R}^n \end{matrix}\right\} \longrightarrow \min_{x\in V}\left\{f(x) : \begin{matrix} \mathcal{A}(x) = b \\ x\in K \end{matrix}\right\} \quad \textbf{(Constrained)}$$

  where f convex, $\mathcal{A}$ linear, K convex

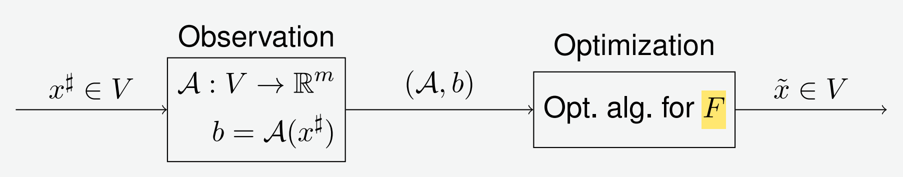


- If no noise in sensing process and no error in optimization algorithm, $\tilde{x} = x^\sharp$
- **Questions**:
  - What if the algorithm receives $\tilde{b} = \mathcal{A}(x^\sharp) + \delta$?
  - What if algorithm only produces a ϵ-optimal and ϵ-feasible solution?
  - What algorithm?
  - Another convex problem?

A sharp penalty formulation

A penalty formulation

- **(Constrained)** $$\min_{x \in V} \left\{ f(x) : \begin{array}{l} \mathcal{A}(x) = b \\ x \in K \end{array} \right\}$$
- Penalty formulation: let $r \asymp \sqrt{k}$ be a penalty parameter
$$F(x) := f(x) + r \left\| \mathcal{A}(x) - b \right\|_1 + 2\,\mathrm{dist}_1(x, K)$$
- Compare: Lasso $\|Ax - b\|_2^2$ vs $\|Ax - b\|_1$



Related: Beck and Teboulle [2009], Tibshirani [1996]

Theorem (Structural)

F is μ-sharp in the ℓ_1 norm (with $\mu > 0$)

$$F(x) - F(x^\sharp) \geq \mu \left\|x - x^\sharp\right\|_1, \qquad \forall x \in V$$

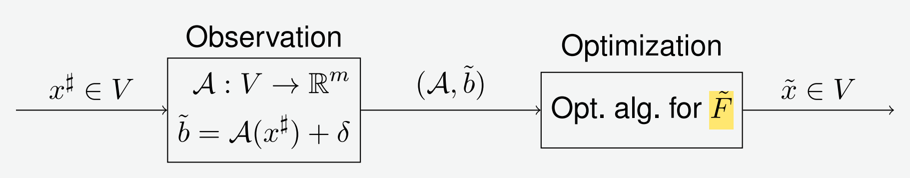
and L-Lipschitz in the ℓ_1 norm with $L \asymp \sqrt{k}$

$$|F(x) - F(y)| \leq L \|x - y\|_1, \qquad \forall x, y.$$

- μ increasing with “RIP constants of $\mathcal{A}$”, in turn depends on sample size
- Sparse recovery: $\mu \asymp 1$ for $m \asymp k \log(n)$

Related: Candes and Tao [2005], Recht et al. [2010]

Robustness of recovery procedure

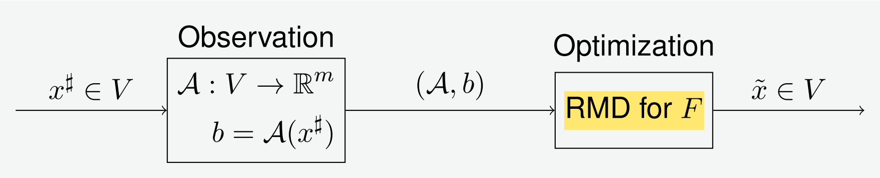


Corollary (Robustness)

Let $\tilde{x}$ be an ϵ minimizer of $\tilde{F}$.

- (to small noise) $\tilde{x}$ satisfies $\left\|\tilde{x} - x^\sharp\right\|_1 \lesssim \frac{\sqrt{k}}{\mu} \left\|\delta\right\|_1 + \frac{\epsilon}{\mu}$
- (to sparse noise) If $\frac{|\mathrm{supp}(\delta)|}{m} \lesssim 1/\sqrt{k}$, then $\left\|\tilde{x} - x^\sharp\right\|_1 \lesssim \frac{\epsilon}{\mu}$

Algorithms for minimizing F



Corollary (Algorithms)

Restarted mirror descent (RMD) algorithm produces an ϵ-optimal solution to F in

$$O\left(\frac{k}{\mu^2}\log(n)\log(\epsilon^{-1})\right)$$

iterations of the mirror descent update.

Related: Polyak [1969], Roulet and d'Aspremont [2017], Yang and Lin [2018], Renegar and Grimmer [2022]

Conclusion

- Abstract statistical signal recovery problem: sparse recovery, covariance estimation, matrix sensing, phase retrieval
- Contributions
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation
  - **Robustness**: to observation error and optimization error
  - **Algorithms**: Restarted Mirror Descent converges linearly

Questions?

References I

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

Candès, E. J., Strohmer, T., and Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274.

Candes, E. J. and Tao, T. (2005). Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215.

Polyak, B. T. (1969). Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29.

Recht, B., Fazel, M., and Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501.

Renegar, J. and Grimmer, B. (2022). A simple nearly optimal restart scheme for speeding up first-order methods. *Foundations of Computational Mathematics*, 22(1):211–256.

Roulet, V. and d'Aspremont, A. (2017). Sharpness, restart and acceleration. *Advances in Neural Information Processing Systems*, 30.

References II

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Yang, T. and Lin, Q. (2018). Rsg: Beating subgradient method without smoothness and strong convexity. *The Journal of Machine Learning Research*, 19(1):236–268.