Sharp exact penalty formulations in signal recovery

Alex L. Wang, Purdue University
Lijun Ding, University of Wisconsin–Madison $\longrightarrow$ Texas A&M

ICIAM 2023

Registration and travel support for this presentation was provided by National Science Foundation Grant DMS-2233032 and the Society for Industrial and Applied Mathematics.

1. Motivation: Sparse recovery

2. Contributions

3. Conclusion

Sparse recovery setup

- **Recovery task**: Recover $x^\sharp \in \mathbb{R}^n$ from $A \in \mathbb{R}^{m \times n}, b = A(x^\sharp)$

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

Sparse recovery setup

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

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

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

Sparse recovery setup

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

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

- **Conceptual approach** $\min_{x\in\mathbb{R}^n} \{|\text{supp}(x)| : Ax = b\}$

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

Sparse recovery setup

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

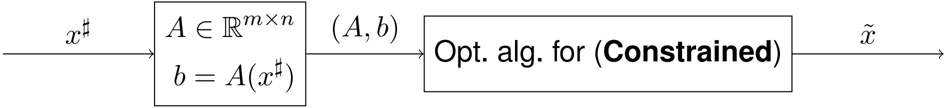
$$\left|\operatorname{supp}(x^\sharp)\right| \leq k \ll n \qquad m \asymp k\log(n)$$

- **Conceptual approach** $\min_{x\in\mathbb{R}^n} \{|\operatorname{supp}(x)| : Ax = b\}$
- **Convex optimization approach**: In this regime, $x^\sharp$ is unique minimizer of

(**Constrained**) $\min_{x\in\mathbb{R}^n} \{\|x\|_1 : Ax = b\}$

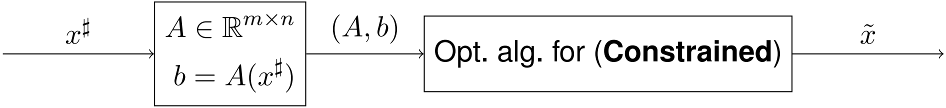
Related: Candes and Tao [2005], Recht et al. [2010], Candès et al. [2013]

Sparse recovery questions



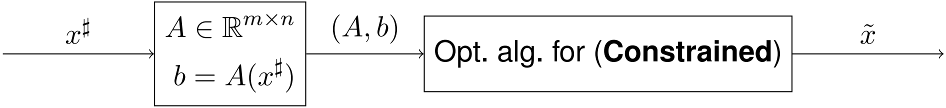
- If no noise in sensing process and no error in optimization algorithm, $\tilde{x} = x^\sharp$

Sparse recovery questions



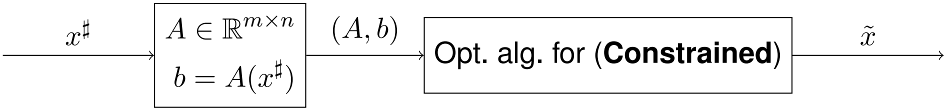
- If no noise in sensing process and no error in optimization algorithm, $\tilde{x} = x^\sharp$
- **Questions**:

Sparse recovery questions



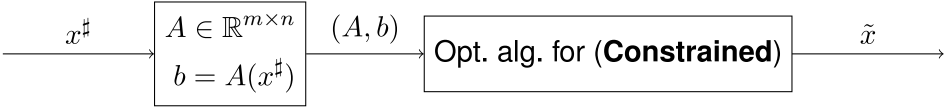
- If no noise in sensing process and no error in optimization algorithm, $\tilde{x} = x^\sharp$
- **Questions**:
  - What if the algorithm receives $\tilde{b} = A(x^\sharp) + \delta$?

Sparse recovery questions



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

Sparse recovery questions



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

Penalty formulation and sharpness

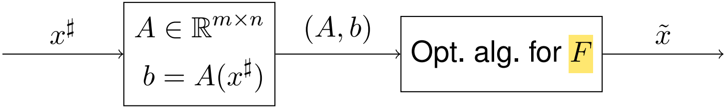
- Penalty formulation: let $r \asymp \sqrt{k}$ and define

$$F(x) := \|x\|_1 + r\,\|Ax - b\|_1$$

Penalty formulation and sharpness

- Penalty formulation: let $r \asymp \sqrt{k}$ and define

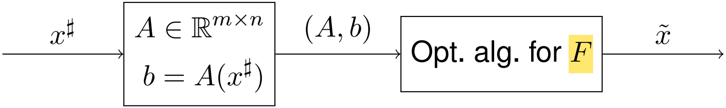
$$F(x) := \|x\|_1 + r\,\|Ax - b\|_1$$



Penalty formulation and sharpness

- Penalty formulation: let $r \asymp \sqrt{k}$ and define

$$F(x) := \|x\|_1 + r\,\|Ax - b\|_1$$



Theorem (Structural)

F is μ-sharp in the ℓ_1 norm with $\mu \asymp 1$

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

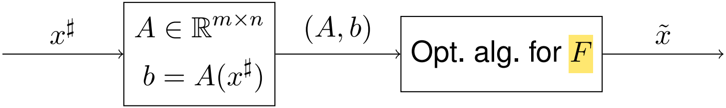
and L-Lipschitz in the ℓ_1 norm with $L \asymp \sqrt{k}$

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

Penalty formulation and sharpness

- Penalty formulation: let $r \asymp \sqrt{k}$ and define

$$F(x) := \|x\|_1 + r\,\|Ax - b\|_1$$



Theorem (Structural)

F is μ-sharp in the ℓ_1 norm with $\mu \asymp 1$

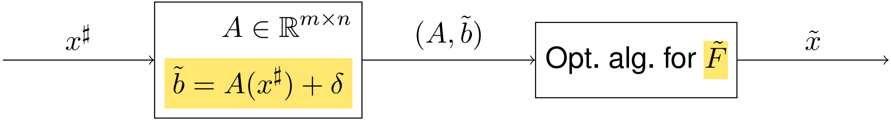
$$F(x) - F(x^\sharp) \geq \mu\,\|x - x^\sharp\|_1\,, \qquad \forall x$$

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

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

- Rearrange proof that $x^\sharp$ optimizes **(Constrained)** with larger RIP constant

Robustness of recovery procedure

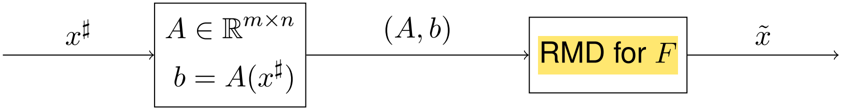


Corollary (Robustness)

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

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

Algorithms for minimizing F



Corollary (Algorithms)

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

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

iterations of the mirror descent update.

Related: Polyak [1969], Roulet et al. [2015], Roulet and d'Aspremont [2017], Yang and Lin [2018], Renegar and Grimmer [2022]

Algorithms for minimizing F

- Suppose we run mirror descent from x_0 for t iterations with step size η and mirror map

$$h(x) \approx \frac{1}{2} \|x - x_0\|_1^2$$

Related: Polyak [1969], Roulet et al. [2015], Roulet and d'Aspremont [2017], Yang and Lin [2018], Renegar and Grimmer [2022]

Algorithms for minimizing F

- Suppose we run mirror descent from x_0 for t iterations with step size η and mirror map

$$h(x) \approx \frac{1}{2}\|x - x_0\|_1^2$$

- Mirror descent guarantee: output y

$$F(y) - F(x^\sharp) \leq \frac{L^2 \eta \ln n}{2} + \frac{D_h(x^\sharp || x_0)}{2\eta t} \approx \frac{L^2 \eta \ln n}{2} + \frac{\left\|x^\sharp - x_0\right\|_1^2}{4\eta t}$$

Related: Polyak [1969], Roulet et al. [2015], Roulet and d'Aspremont [2017], Yang and Lin [2018], Renegar and Grimmer [2022]

Algorithms for minimizing F

- Suppose we run mirror descent from x_0 for t iterations with step size η and mirror map

$$h(x) \approx \frac{1}{2} \|x - x_0\|_1^2$$

- Mirror descent guarantee: output y

$$F(y) - F(x^\sharp) \le \frac{L^2 \eta \ln n}{2} + \frac{D_h(x^\sharp || x_0)}{2\eta t} \approx \frac{L^2 \eta \ln n}{2} + \frac{\left\|x^\sharp - x_0\right\|_1^2}{4\eta t}$$

- Optimizing in η and applying sharpness $\longrightarrow$

$$F(y) - F(x^\sharp) \le \frac{1}{2} \left(F(x_0) - F(x^\sharp)\right)$$

after $\asymp \frac{L^2}{\mu^2} \ln n$ iterations

Related: Polyak [1969], Roulet et al. [2015], Roulet and d'Aspremont [2017], Yang and Lin [2018], Renegar and Grimmer [2022]

Conclusion

- Sparse recovery

Conclusion

- Sparse recovery
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation

Conclusion

- Sparse recovery
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation
  - **Robustness**: to observation error and optimization error

Conclusion

- Sparse recovery
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation
  - **Robustness**: to observation error and optimization error
  - **Algorithms**: Restarted Mirror Descent converges linearly

Conclusion

- Sparse recovery
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation
  - **Robustness**: to observation error and optimization error
  - **Algorithms**: Restarted Mirror Descent converges linearly
- Similar guarantees for phase retrieval, low-rank matrix sensing, covariance estimation

Conclusion

- Sparse recovery
  - **Structural**: ℓ_1 sharp and Lipschitz penalty formulation
  - **Robustness**: to observation error and optimization error
  - **Algorithms**: Restarted Mirror Descent converges linearly
- Similar guarantees for phase retrieval, low-rank matrix sensing, covariance estimation

Thank you! Questions?

`https://arxiv.org/abs/2307.06873`

References I

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., Boumal, N., and d'Aspremont, A. (2015). Renegar's condition number, sharpness and compressed sensing performance. *arXiv preprint arXiv:1506.03295*.

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

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.