**Sparse High-Dim Linear Regression**

- Often $p > n$
- $\beta_j \neq 0$ means the $j$th variable is relevant
- Most of entries of $\beta$ are zeros

**SLOPE Problem and Properties**

Bogdan et al. (2015) proposed SLOPE problem as recovering:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \| y - X \beta \|_2^2 + \lambda_1 \| \beta \|_1$$

where $J_\lambda(b) = \lambda_1 |b|_1 + \cdots + \lambda_p |b|_p$

having $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and $|b|_p \geq \cdots \geq |b|_1$ are the order statistics. Clearly, when $\lambda_1 = \cdots = \lambda_p$, SLOPE reduces to LASSO.

**Computing SLOPE Solution**

Denote $\text{prox}_\phi(y, \lambda) = \arg \min_b \frac{1}{2} \| y - b \|_2^2 + \lambda \phi(b)$.

We may solve the SLOPE problem by

- Subgradient method: $\beta^{t+1} = \beta^{(t)} - s_t \nabla \phi^{(t)}$
  where $\nabla \phi^{(t)}$ is a subgradient of objective function at $\beta^{(t)}$.
- ISTA: Iterative Shrinkage Thresholding Algorithm, $\beta^{(t+1)} = \text{prox}_{\lambda \phi}(\beta^{(t)} + \lambda X^T (y - X \beta^{(t)}))$
- FISTA: Fast ISTA, with $s_{t+1} = (1 + \sqrt{1 + 4s_t^2})/2$
  $\beta^{(t+1)} = \text{prox}_{\lambda \phi}(\beta^{(t)} + s_{t+1} \lambda X^T (y - X \beta^{(t)}))$
- AMP: Approximate Message Passing [Donoho-Maleki-Montanari '10],
  $\beta^{t+1} = \text{prox}_{\lambda \phi}(X^T r^{t} + \beta_t)$
  $r^{t+1} = y - X \beta^{t+1} + \frac{1}{n} \nabla \text{prox}_{\lambda \phi}(X^T r^{t} + \beta_t)$.

**Proximal Operators**

- For any function $\phi$, $\text{prox}_\phi(a, \lambda)$ is defined as $\text{prox}_\phi(a, \lambda) = \arg \min_b \frac{1}{2} \| a - b \|_2^2 + \lambda \phi(b)$
  and $\nabla \text{prox}_\phi$ is divergence of the proximal operator
- There exists an algorithm to compute the proximal operator when $h = J_\lambda$ [Bogdan-Sabatt-Su-Candes '15]
- For SLOPE, $\nabla \text{prox}_{\lambda \phi}(b) = \| \text{prox}_{\lambda \phi}(b) \|_2$ where $\| \text{prox}_{\lambda \phi}(b) \|_2$ counts the unique non-zero magnitudes in $b$.

**Main Results of SLOPE**

**Theorem 1** Under conditions on $\lambda$, the state evolution recursion with calibration defined above, has a unique fixed point to which the convergence monotonic in $t$, for any initial condition.

**Calibration** between $\lambda$ and $\alpha$:

$$\lambda = \alpha \tau \left( 1 - \lim_{t \to \infty} \frac{1}{\delta} | \text{prox}_{\lambda \phi}(b + \tau Z) - b |^2 \right)$$

**Theorems**

- **Theorem 2** Under some assumptions, $\lim_{t \to \infty} \frac{1}{t} \| \hat{\beta} - \beta \|_2^2 = c_t$, where $\lim_{t \to \infty} c_t = 0$.
- **Theorem 3** Under some assumptions, for any uniformly pseudo-Lipschitz sequence of functions $\phi_t$ and for $Z \sim N(0, I_p)$,
  $$\lim_{t \to \infty} \frac{1}{t} \| \phi_t(\hat{\beta} - \beta) \|_2^2 = \delta (\tau^2 - \sigma^2)$$

**Corollary 3.1** Under some assumptions,
  $$\lim_{t \to \infty} \frac{1}{t} \| \hat{\beta} - \beta \|_2^2 = \delta (\tau^2 - \sigma^2)$$

**Challenge**

- For large enough $\epsilon = |\text{supp}(\beta)|/p$ or small enough $\delta = n/p$, LASSO suffers from Donoho-Tanner phase transition: TPP is bounded away from 1 [Donoho-Tanner '09; Su-Bogdan-Candes '17 (image source)].

**Inference:** **TPP, FDP & MSE**

- However, SLOPE overcomes the phase transition. Specifically we can characterize one of the SLOPE path as a M"obius transformation: for TPP= $u$, the minimum FDP is at most $\frac{\log(p)}{\log(u)}$ for some constants $a, b, c, d$.

- Figure 1: Red dot: LASSO; Blue dot: SLOPE; Black solid line: LASSO trade-off; Red dashed line: SLOPE trade-off.

- In addition, fixing the signal prior and under some assumptions, we show that switching from LASSO to SLOPE gives better paths in the sense of achieving smaller FDP, larger TPP and smaller mean squared error at the same time.