This document combines a number of threads for model selection in our problem (and other related optimization problems). First, we consider estimating the degrees of freedom. Then we give two results which can be used to select models.

1 Degrees of freedom for Gen-Gen Lasso

Suppose we are minimizing the negative log likelihood from a natural exponential family subject to the generalized lasso penalty. That is, given $Y_i \sim p(y \mid \eta_i(\tau))$ for $\tau \in \mathbb{R}^p$, $i = 1, \ldots, n$, our goal is to solve

$$\widehat{\tau} = \operatorname*{argmin}_{\tau} \sum_{i=1}^{n} \phi(\eta_i(\tau)) - \langle y_i,\ \eta_i(\tau) \rangle + \lambda \left\| D\tau \right\|_1, \tag{1}$$

where $D \in \mathbb{R}^{q \times p}$.

This generalizes the various standard models. For example, ℓ_1-trend filtering has $\phi(x) = x^2/2$, $\eta_i(x) = x_i$, and D the first-order discrete difference operator. Variance estimation in our context has $\phi(x) = x$, $y_i = z_i^2$ (where $z_i \sim N(0, \exp(-h))$ and $\eta_i(x) = -e^{-\tau_i}$. Logistic loss also falls into this category.

We first consider the case that $\eta_i(\tau) = \eta_i(\tau_i)$, i.e., $\tau \in \mathbb{R}^n$ and $p(y_i \mid \tau) = p(y_i \mid \tau_i)$.

Theorem 0.1. *The divergence of $\widehat{\tau}(y)$ is given by*

$$\operatorname{tr}\left(D\widehat{\tau}(y)\right) = -P_{\mathcal{N}(D)} \left(P_{\mathcal{N}(D)} \operatorname{diag}\left(\frac{d^2}{d\tau_i^2} \ell|_{y_i, \widehat{\eta}_i} \right) P_{\mathcal{N}(D)} \right)^{\dagger} P_{\mathcal{N}(D)} \operatorname{diag}\left(\frac{d^2}{d\tau_i dy_i} \ell|_{y_i, \widehat{\eta}_i} \right), \tag{2}$$

where $\ell = \ell(y, \eta(\tau_i)) = \sum_{i=1}^n \phi(\eta_i(\tau_i)) - \langle y_i,\ \eta_i(\tau_i) \rangle$, $\widehat{\eta}_i = \eta_i(\widehat{\tau}_i)$, and

$$P_{\mathcal{N}(D)} = I_n - D_S^{\top}(D_S^{\top} D_S)^{\dagger} D_S \tag{3}$$

is the projection onto the null-space of D_S, where $S = \{j \in [q] : D\widehat{\tau} = 0\}$, and the notation D_S means the rows of D whose indices are in S.

Special cases:

1. If we are interested in the natural exponential family (that is $\eta_i(\tau_i) = \tau_i$), then $\frac{d^2}{d\tau_i dy_i}\ell = -1$ and $\frac{d^2}{d\tau_i^2}\phi_i(\eta_i(\tau_i)) = \operatorname{Var}[y_i]$.
2. For Gaussian likelihood, $\frac{d^2}{d\tau_i^2}\ell = 1$ and $\frac{d^2}{d\tau_i dy_i}\ell = -1$, so the divergence is the dimension of $\mathcal{N}(D_S)$ as shown in Tibshirani and Taylor [1].
3. For logistic loss, $\frac{d^2}{d\tau_i^2}\ell = e^{\tau_i}/(1+e^{\tau_i})^2$ and $\frac{d^2}{d\tau_i dy_i}\ell = -1$.
4. For the case of variance estimation, we have $\frac{d^2}{d\tau_i^2}\ell = y_i^2 e^{-h}$ and $\frac{d^2}{d\tau_i dy_i}\ell = -e^{-h}$.

Now we generalize to the regression setting. Define $\mu_i = x_i^{\top}\beta$. Furthermore, write

$$\widehat{\beta} = \operatorname*{argmin}_{\tau} \sum_{i=1}^{n} \phi(\mu_i(\beta)) - \langle y_i,\ \mu_i(\beta) \rangle + \lambda \left\| D\beta \right\|_1. \tag{4}$$

Theorem 0.2. *The divergence of $\widehat{\mu}(y)$ is given by*

$$\operatorname{tr}\left(D\widehat{\mu}(y)\right) = -X_P \left(X_P^{\top} \operatorname{diag}\left(\frac{d^2}{d\mu_i^2} \ell|_{y_i, \widehat{\mu}_i} \right) X_P \right)^{\dagger} X_P^{\top} \operatorname{diag}\left(\frac{d^2}{d\mu_i dy_i} \ell|_{y_i, \widehat{\mu}_i} \right), \tag{5}$$

where $\ell = \ell(y, \eta(\mu_i)) = \sum_{i=1}^{n} \phi(\eta_i(\mu_i)) - \langle y_i,\ \mu_i \rangle$, $\widehat{\eta}_i = \eta_i(x_i^\top \widehat{\beta})$, *and*

$$X_P = (I_n - D_S^\top (D_S^\top D_S)^\dagger D_S) X \tag{6}$$

is the projection of X *onto the null-space of* D_S.

Proof. This follows mainly from Theorem 2 in [2], though the conditions are non-trivial. To be expanded. □

References

[1] Tibshirani, R. J., and Taylor, J. (2012), "Degrees of freedom in lasso problems," *Annals of Statistics*, **40**, 1198–1232, arXiv:1111.0653.

[2] Vaiter, S., Deledalle, C., Fadili, J., Peyré, G., and Dossal, C. (2017), "The degrees of freedom of partly smooth regularizers," *Annals of the Institute of Statistical Mathematics*, **69**, 791–832.