计量经济学中的凸优化

史震涛 (with 高展)

香港中文大学

广州计量经济学高级研讨会
2017年11月11日

Convex Optimization

- Big data vs. big model
- General optimization vs. convex optimization

Computing Environment

- Convex solvers: `MOSEK`, `Gurobi`
- High-level: `Matlab+CVX`:
  - Su, Shi and Phillips (2016), Shi (2016b)
- Lower-level: `Matlab+TFOCS`

- High-level: `R+cvxr` (yet to come)
- Lower-level: `R+Rmosek`
  - Koenker and Mizera (2014a)
  - Gao and Shi (2017)

Computing Environment

- Convex solvers: `MOSEK`, `Gurobi`
- High-level: `Matlab+CVX`:
  - Su, Shi and Phillips (2016), Shi (2016b)
- Lower-level: `Matlab+TFOCS`

- High-level: `R+cvxr` (yet to come)
- Lower-level: `R+Rmosek`
  - Koenker and Mizera (2014a)
  - Gao and Shi (2017)

Lasso

- Tibshirani (1996)

$$\min_{\beta} N^{-1} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

```
begin_cvx
        variable b(p);
        objective = sum_square( y - X * b )/n ...
                + lambda * norm(b, 1);
        minimize(objective)
end_cvx
```

GMM-Lasso

- Caner (2009), Shi (2016a)

$$\left\| n^{-1} W_n^{1/2} Z' (y - X\beta) \right\|_2^2 + \lambda \|\beta\|_1 .$$

- Rewrite as

$$\left\| \tilde{y} - \tilde{X}\beta \right\|_2^2 + \lambda \|\beta\|_1 ,$$

  where $\tilde{y} = n^{-1} W_n^{1/2} Z' y$ and $\tilde{X} = n^{-1} W_n^{1/2} Z' X$.
- Shi (2016a) implements it by modifying `LARS`.

Classifier-Lasso

- Bonhomme and Manresa (2015)
- Su, Shi and Phillips (2016)
  - Penalized least squares (PLS)

$$\min_{\boldsymbol{\beta},(\alpha_k)_{k=1}^K} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-x_{it}'\beta_i\right)^2+\frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i-\alpha_k\|_2 .$$

Classifier-Lasso (continue)

- Iterative algorithm.
- In the k-th sub-step of the r-th iteration, we choose $(\boldsymbol{\beta}, \alpha_{\tilde{k}})$ to minimize

$$\min_{\boldsymbol{\beta}, \alpha_{\tilde{k}}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_i)^2 + \frac{\lambda}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_{\tilde{k}}\|_2 \gamma_i$$

where $\gamma_i = \prod_{k=1}^{\tilde{k}-1} \left\| \hat{\beta}_i^{(r,k)} - \hat{\alpha}_k^{(r)} \right\|_2 \prod_{k=\tilde{k}+1}^{K} \left\| \hat{\beta}_i^{(r-1,k)} - \hat{\alpha}_k^{(r-1)} \right\|_2 .$

- Consistent initial estimator of β_i when T is large.
- Extend to nonlinear MLE and GMM.
  - PPML and PGMM

Classifier-Lasso (continue)

- Iterative algorithm.
- In the k-th sub-step of the r-th iteration, we choose $(\boldsymbol{\beta}, \alpha_{\tilde{k}})$ to minimize

$$\min_{\boldsymbol{\beta},\alpha_{\tilde{k}}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_i)^2 + \frac{\lambda}{N}\sum_{i=1}^{N} \|\beta_i - \alpha_{\tilde{k}}\|_2 \gamma_i$$

where $\gamma_i = \prod_{k=1}^{\tilde{k}-1} \left\| \hat{\beta}_i^{(r,k)} - \hat{\alpha}_k^{(r)} \right\|_2 \prod_{k=\tilde{k}+1}^{K} \left\| \hat{\beta}_i^{(r-1,k)} - \hat{\alpha}_k^{(r-1)} \right\|_2 .$

- Consistent initial estimator of β_i when T is large.
- Extend to nonlinear MLE and GMM.
  - PPML and PGMM

Dantzig Selector

- Candes and Tao (2007)

$$\min_{\beta} \|\beta\|_1 \ \text{ s.t. } \ \|X'(y - X\beta)\|_\infty \le \lambda.$$

```
begin_cvx
        minimize( norm(b, 1) )
        subject to
        X'(y - X * b) <= lambda
        -X'(y - X * b) <= lambda
end_cvx
```

Empirical Likelihood

- Unconditional moment conditions $\mathbb{E}\left[g\left(Z_i,\beta\right)\right]=\mathbf{0}_m$.
- GMM and EL.
- EL (Qin and Lawless, 1993; Kitamura, 1997) solves

$$\min_{\beta\in\mathcal{B},\pi\in\Delta_n} -\sum_{i=1}^{n}\log\pi_i \quad \text{s.t.} \quad \sum_{i=1}^{n}\pi_i g\left(Z_i,\beta\right)=\mathbf{0}_m$$

  where $\Delta_n=\left\{\pi\in[0,1]^n : \sum_{i=1}^{n}\pi_i=1\right\}$ is the n-dimensional probability simplex.
- Two loops

$$\min_{\beta\in\mathcal{B}}\min_{\pi\in\Delta_n(\beta)} -\sum_{i=1}^{n}\log\pi_i$$

  where $\Delta_n\left(\beta\right)=\left\{\pi\in\Delta_n : \sum_{i=1}^{n}\pi g\left(Z_i,\beta\right)=\mathbf{0}_m\right\}$.

Empirical Likelihood

- Unconditional moment conditions $\mathbb{E}\left[g\left(Z_i,\beta\right)\right]=\mathbf{0}_m$.
- GMM and EL.
- EL (Qin and Lawless, 1993; Kitamura, 1997) solves

$$\min_{\beta\in\mathcal{B},\pi\in\Delta_n} -\sum_{i=1}^{n}\log\pi_i \quad \text{s.t.} \quad \sum_{i=1}^{n}\pi_i g\left(Z_i,\beta\right)=\mathbf{0}_m$$

where $\Delta_n=\left\{\pi\in[0,1]^n:\sum_{i=1}^n\pi_i=1\right\}$ is the n-dimensional probability simplex.

- Two loops

$$\min_{\beta\in\mathcal{B}}\min_{\pi\in\Delta_n(\beta)} -\sum_{i=1}^{n}\log\pi_i$$

where $\Delta_n\left(\beta\right)=\left\{\pi\in\Delta_n:\sum_{i=1}^n\pi g\left(Z_i,\beta\right)=\mathbf{0}_m\right\}$.

Empirical Likelihood

- Unconditional moment conditions $\mathbb{E}\left[g\left(Z_i, \beta\right)\right] = \mathbf{0}_m$.
- GMM and EL.
- EL (Qin and Lawless, 1993; Kitamura, 1997) solves

$$\min_{\beta \in \mathcal{B}, \pi \in \Delta_n} -\sum_{i=1}^{n} \log \pi_i \quad \text{s.t.} \quad \sum_{i=1}^{n} \pi_i g\left(Z_i, \beta\right) = \mathbf{0}_m$$

  where $\Delta_n = \left\{\pi \in [0,1]^n : \sum_{i=1}^{n} \pi_i = 1\right\}$ is the n-dimensional probability simplex.
- Two loops

$$\min_{\beta \in \mathcal{B}} \min_{\pi \in \Delta_n(\beta)} -\sum_{i=1}^{n} \log \pi_i$$

  where $\Delta_n\left(\beta\right) = \left\{\pi \in \Delta_n : \sum_{i=1}^{n} \pi g\left(Z_i, \beta\right) = \mathbf{0}_m\right\}$.

Relaxed Empirical Likelihood

- Shi (2016b)

$$\min_{\beta\in\mathcal{B},\pi\in\Delta_n} -\sum_{i=1}^{n}\log\pi_i \quad \text{s.t.} \quad \left\| \sum_{i=1}^{n}\pi_i g(Z_i,\beta) \right\|_\infty \le \lambda.$$

- Inner loop convex, outer loop low-dimensional.

REL: Dual Problem

- Alternatively, the primal problem of the inner loop can be written as

$$\min_{\pi \in \Delta_n} - \sum_{i=1}^{n} \log \left(1 + \lambda' g\left(Z_i, \beta\right)\right) + \lambda \left\|\gamma\right\|_1$$

  where γ is the Lagrangian multiplier.
- Carrasco and Kotchoni (2017), Chang, Tang and Wu (2017)

More Examples

- Belloni, Chernozhukov and Wang (2011)
  - square-root Lasso
- Moon and Weidner (no paper online yet)
  - nuclear norm for Bai (2009)'s interactive fixed-effect model
- Koenker and Mizera (2014b)
  - Kiefer-Wolfowitz Bayes rule

Conclusion

- Convex programming is convenient and reliable
- Theoretical properties