Big Data Sketching with Model Mismatch



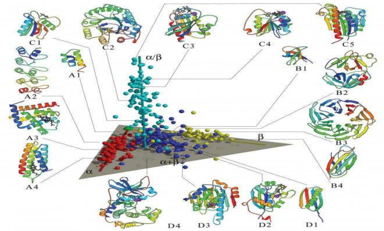

Sundeep Chepuri **Yu Zhang** **Geert Leus** **Georgios Giannakis**

ASILOMAR 2015, Pacific Grove, USA

TU Delft Delft University of Technology

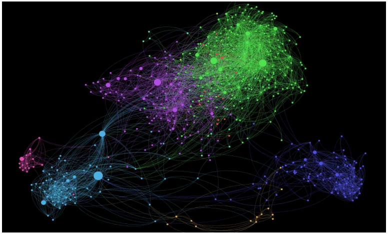
University of Minnesota

Power networks, grid analytics



Biological networks

Oil and gas field exploration



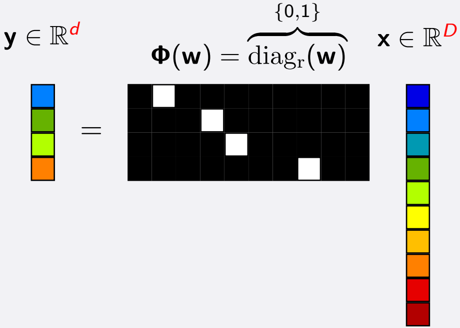
Internet, social media

Massive data, but limited computational capacity

Sketching or Censoring

- Sketching or Censoring — tool for data reduction.
- Why sketching?
  - Reduce (inferential) processing overhead
  - Quick rough answer
- How is sketching done?
  - Random sampling
    [Drineas-Mahoney-Muthukrishnan-2006], [Strohmer-Vershynin-2009]
  - Design of experiments (censoring—distributed setup)
    [Rago-Willett-Bar-Shalom-1996], [Msechu-Giannakis-2012], [Berberidis-Kekatos-Giannakis-2015]

Sparse sampling for sketching



What is sparse sampling?

Design $\mathbf{w} \in \{0,1\}^D$ to select the most "informative" d ($\ll D$) samples

$\mathrm{diag_r}(\cdot)$ - diagonal matrix with the argument on its diagonal but with the zero rows removed.

Linear regression — model mismatch

- Observations follow

$$x_m = \bar{\mathbf{a}}_m^T \boldsymbol{\theta} + n_m, \ m = 1, 2, \ldots, D$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ Unknown parameter
  - n_m i.i.d. zero-mean unit-variance Gaussian noise

- Regressors are known up to a bounded uncertainty

$$\bar{\mathbf{a}}_m = \underbrace{\mathbf{a}_m}_{\text{known}} + \underbrace{\mathbf{p}_m}_{\text{unknown}, \|\mathbf{p}_m\|_2 \leq \eta}$$

Problem statement

Given $\{x_m\}$, $\{\mathbf{a}_m\}$, and η,

(a) design $\mathbf{w}$ to censor less-informative samples

(b) estimate $\boldsymbol{\theta}$ that performs well for any allowed $\{\mathbf{p}_m\}$

Optimization problem

- Censored robust least squares (min. the worst-case residual)

$$\min_{\mathbf{w}\in\mathcal{W},\boldsymbol{\theta}} \max_{\|\mathbf{p}_m\|_2\leq\eta, m=1,2,\ldots,D} \sum_{m=1}^{D} w_m \left(x_m - (\mathbf{a}_m + \mathbf{p}_m)^T\boldsymbol{\theta}\right)^2$$

$\mathcal{W} = \{\mathbf{w} \in \{0,1\}^D \mid \|\mathbf{w}\|_0 = d\}$.

- Min-max problem is equivalent to the min. problem

$$\min_{\mathbf{w}\in\mathcal{W},\boldsymbol{\theta}} \overbrace{\sum_{m=1}^{D} w_m \left(|x_m - \mathbf{a}_m^T\boldsymbol{\theta}| + \eta\|\boldsymbol{\theta}\|_2\right)^2}^{\text{worst-case residual}}$$

- Problem simplifies to censored least-squares for $\eta = 0$

Proposed solver

Optimization problem

$$\min_{\mathbf{w}\in\mathcal{W},\boldsymbol{\theta}} \sum_{m=1}^{D} w_m \overbrace{\left(|x_m - \mathbf{a}_m^T\boldsymbol{\theta}| + \eta\|\boldsymbol{\theta}\|_2\right)^2}^{r_m^2(\boldsymbol{\theta})}$$

- For fixed $\{w_m\}$, it is robust least squares
  [Ghaoui-Lebret-1997], [Chandrashekaran-Golub-Gu-Sayed-1998]
- For large values of η, $\boldsymbol{\theta}^\star = \mathbf{0}$
- $\{w_m\}$ are Boolean

Proposed solver

- Nonconvex Boolean optimization problem

$$\min_{\mathbf{w}\in\mathcal{W},\boldsymbol{\theta}} \sum_{m=1}^{D} w_m \overbrace{\left(|x_m - \mathbf{a}_m^T\boldsymbol{\theta}| + \eta\|\boldsymbol{\theta}\|_2\right)^2}^{r_m^2(\boldsymbol{\theta})} \Leftrightarrow \min_{\boldsymbol{\theta}} \sum_{m=1}^{d} r_{[m]}^2(\boldsymbol{\theta})$$

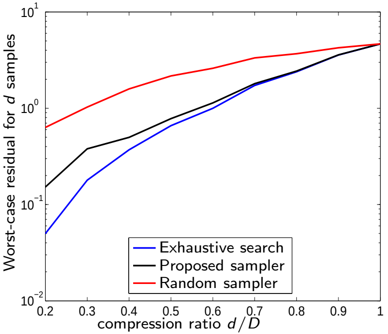
$r_{[m]}^2(\boldsymbol{\theta})$ are squared regularized residuals in ascending order

- simplifies to simple low-complexity problems:

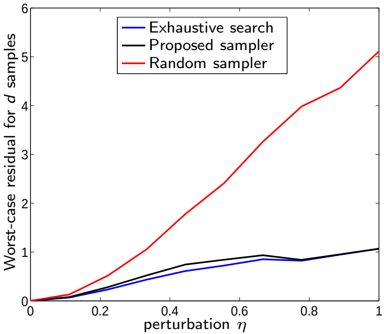
Alternatively update $\mathbf{w}$ and $\boldsymbol{\theta}$

- For a given $\boldsymbol{\theta}$, the optimal $\mathbf{w}$ is obtained by ordering the regularized residuals.
- For a given $\mathbf{w}$, $\boldsymbol{\theta}$ is obtained by solving the reduced-order ($d \ll D$) regularized least-squares
  - convex/SOCP; or even, first-order methods

- Random (Gaussian) regression matrix
- $D = 10,\ p = 2$



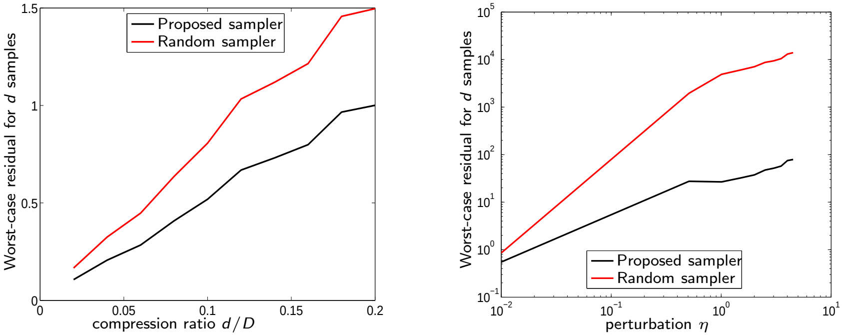
$\eta = 0.5$



50% compression

$D = 5000$—synthetic

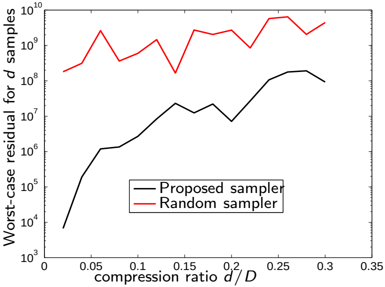
- Random (Gaussian) regression matrix
- $D = 5000$, $p = 10$



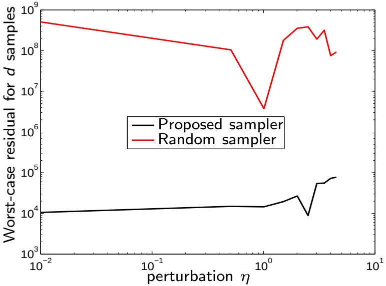
$\eta = 0.01$ | 10% compression

Real dataset — protein (tertiary) structure modeling

- Entries of the regression matrix contain structure revealing parameters obtained via experiments (hence are perturbed/noisy)
- Observations are distance to native proteins.
- $D = 45730,\ p = 9$



$\eta = 0.01$



1% compression

Conclusions and future directions

- Design censoring scheme for linear regression
  - In presence of bounded uncertainties
  - Data dependent by nature
- Streaming data (not batch)
  - online algorithms (e.g., recursive least squares-like) need to be devised
- Sketching with model mismatch
  - Correlated observations, clustering, and classification

Thank You!!

Selected references

- Drineas, P., Mahoney, M. W., Muthukrishnan, S. (2006, January). Sampling algorithms for ℓ_2 regression and applications. In Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (pp. 1127-1136). Society for Industrial and Applied Mathematics.
- Strohmer, T., Vershynin, R. (2009). A randomized Kaczmarz algorithm with exponential convergence. Journal of Fourier Analysis and Applications, 15(2), 262-278.
- Rago, C., Willett, P., Bar-Shalom, Y. (1996). Censoring sensors: A low-communication-rate scheme for distributed detection. Aerospace and Electronic Systems, IEEE Transactions on, 32(2), 554-568.
- Berberidis, D., Kekatos, V., Giannakis, G. B. (2015). Online Censoring for Large-Scale Regressions with Application to Streaming Big Data. arXiv preprint arXiv:1507.07536.
- Msechu, E. J., Giannakis, G. B. (2012). Sensor-centric data reduction for estimation with WSNs via censoring and quantization. Signal Processing, IEEE Transactions on, 60(1), 400-414.
- Chandrasekaran, S., Golub, G. H., Gu, M., Sayed, A. H. (1998). Parameter estimation in the presence of bounded data uncertainties. SIAM Journal on Matrix Analysis and Applications, 19(1), 235-252.
- El Ghaoui, L., Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. SIAM Journal on Matrix Analysis and Applications, 18(4), 1035-1064.