Randomized Least Squares Regression: Combining Model- and Algorithm-Induced Uncertainties

Ilse C.F. Ipsen

Joint work with: Jocelyn T. Chi

North Carolina State University
Raleigh, NC, USA

Least Squares/Regression Models

Given: Design matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ with $\text{rank}(\boldsymbol{X}) = p$

❶ Gaussian linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$$

Unknown parameter vector $\boldsymbol{\beta}_0$

❷ Least squares problem

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2$$

Unique maximum likelihood estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$

❸ Randomized (row compression) algorithm

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{S}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})\|_2$$

Minimal norm solution $\tilde{\boldsymbol{\beta}} = (\boldsymbol{S}\boldsymbol{X})^\dagger (\boldsymbol{S}\boldsymbol{y})$

Objective

Determine combined mean and variance of $\tilde{\boldsymbol{\beta}}$ w.r.t.

- Gaussian linear model $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$
- Randomized row compression $\boldsymbol{S}$

Inspiration

Ping Ma, Michael Mahoney, Bin Yu
A statistical perspective on algorithmic leveraging
J. Mach. Learn. Res., vol. 16, pp 861-911 (2015)

Overview

Existing work
Examples
Structural perturbation bounds
Model-induced uncertainty, conditioned on algorithm-induced uncertainty
Combined model-induced and algorithm-induced uncertainty
Summary

Existing Work

(Numerical) row-compression methods for least squares

Drineas, Mahoney, Muthukrishnan 2006
Zhou, Lafferty, Wasserman 2007
Boutsidis, Drineas 2009
Drineas, Mahoney, Muthukrishnan, Sarlós 2011
Meng, Saunders, Mahoney 2014
Bartels, Hennig 2016
Becker, Kawas, Petrick, Ramamurthy 2017

Statistical properties

Ma, Mahoney, Yu 2015
Raskutti, Mahoney 2016
Ahfock, Astle, Richardson 2017
Thanei, Heinze, Meinshausen 2017
Lopes, Wang, Mahoney 2018

Examples

Model-Induced Uncertainty

Gaussian linear model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$

Design matrix:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{X}^\dagger = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Left inverse: $\boldsymbol{X}^\dagger \boldsymbol{X} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Maximum likelihood estimator: $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\dagger \boldsymbol{y}$

Unbiased estimator: $\mathbb{E}_{\boldsymbol{y}}[\hat{\boldsymbol{\beta}}] = \boldsymbol{X}^\dagger \, \mathbb{E}_{\boldsymbol{y}}[\boldsymbol{y}] = \boldsymbol{X}^\dagger \boldsymbol{X} \boldsymbol{\beta}_0 = \boldsymbol{\beta}_0$

Variance: $\mathbb{V}\mathrm{ar}_{\boldsymbol{y}}[\hat{\boldsymbol{\beta}}] = \boldsymbol{X}^\dagger \, \mathbb{V}\mathrm{ar}_{\boldsymbol{y}}[\boldsymbol{y}] \left(\boldsymbol{X}^\dagger\right)^T = \sigma^2 \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$

Model-Induced Uncertainty
Conditioned on Algorithm-Induced Uncertainty (1)

$\min_\beta \|\boldsymbol{S}\boldsymbol{X}\beta - \boldsymbol{S}\boldsymbol{y}\|_2$ has solution $\tilde{\beta} = (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\,\boldsymbol{y}, \quad \mathbb{E}_{\boldsymbol{y}}[\boldsymbol{y}] = \boldsymbol{X}\beta_0$

Sketching preserves rank: $\operatorname{rank}(\boldsymbol{S}\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{X})$

$$\boldsymbol{S}\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (\boldsymbol{S}\boldsymbol{X})^\dagger$$

Unbiased estimator:

$$\mathbb{E}_{\boldsymbol{y}}\left[\tilde{\beta} \mid \boldsymbol{S}\right] = (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\, \mathbb{E}_{\boldsymbol{y}}[\boldsymbol{y}] = (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\boldsymbol{X}\beta_0 = \beta_0$$

Variance has increased:

$$\mathbb{V}\mathrm{ar}_{\boldsymbol{y}}\left[\tilde{\beta} \mid \boldsymbol{S}\right] = \sigma^2 (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S} \left((\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\right)^T = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \succcurlyeq \sigma^2 \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$$

Model-Induced Uncertainty
Conditioned on Algorithm-Induced Uncertainty (2)

$\min_\beta \|S(X\beta - y)\|_2$ has solution $\tilde{\beta} = (SX)^\dagger S y, \quad \mathbb{E}_y[y] = X\beta_0$

Sketching causes loss of rank: $\mathrm{rank}(SX) < \mathrm{rank}(X)$

$$SX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = (SX)^\dagger$$

Biased estimator:

$$\mathbb{E}_y\left[\tilde{\beta} \,\middle|\, S\right] = (SX)^\dagger SX \beta_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \beta_0 \neq \beta_0$$

Variance is singular:

$$\mathbb{V}\mathrm{ar}_y\left[\tilde{\beta} \,\middle|\, S\right] = \sigma^2 (SX)^\dagger S \left((SX)^\dagger S\right)^T = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Model-Induced Uncertainty
Conditioned on Algorithm-Induced Uncertainty (3)

$\min_{\beta} \| \boldsymbol{S}(\boldsymbol{X}\beta - \boldsymbol{y}) \|_2$ has solution $\tilde{\beta} = (\boldsymbol{S}\boldsymbol{X})^{\dagger} \boldsymbol{S}\boldsymbol{y}, \quad \mathbb{E}_{\boldsymbol{y}}[\boldsymbol{y}] = \boldsymbol{X}\beta_0$

Summary: Model-induced uncertainty conditioned on $\boldsymbol{S}$

- Sketching preserves rank: $\mathrm{rank}(\boldsymbol{S}\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{X})$

    Left inverse: $(\boldsymbol{S}\boldsymbol{X})^{\dagger} = (\boldsymbol{X}^T \boldsymbol{S}^T \boldsymbol{S}\boldsymbol{X})^{-1} (\boldsymbol{S}\boldsymbol{X})^T$

    $\tilde{\beta}$ is an unbiased estimator: $\mathbb{E}_{\boldsymbol{y}}[\tilde{\beta} \mid \boldsymbol{S}] = \beta_0$

$\min_{\beta} \|\boldsymbol{S}(\boldsymbol{X}\beta - \boldsymbol{y})\|_2$ has solution $\tilde{\beta} = (\boldsymbol{S}\boldsymbol{X})^{\dagger}\boldsymbol{S}\,\boldsymbol{y}, \quad \mathbb{E}_{\boldsymbol{y}}[\boldsymbol{y}] = \boldsymbol{X}\beta_0$

Summary: Model-induced uncertainty conditioned on $\boldsymbol{S}$

- Sketching preserves rank: $\mathrm{rank}(\boldsymbol{S}\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{X})$

    Left inverse: $(\boldsymbol{S}\boldsymbol{X})^{\dagger} = (\boldsymbol{X}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{X})^{-1}(\boldsymbol{S}\boldsymbol{X})^T$

    $\tilde{\beta}$ is an unbiased estimator: $\mathbb{E}_{\boldsymbol{y}}[\tilde{\beta}\,\big|\,\boldsymbol{S}] = \beta_0$

- Sketching causes loss of rank: $\mathrm{rank}(\boldsymbol{S}\boldsymbol{X}) < \mathrm{rank}(\boldsymbol{X})$

    No left inverse: $(\boldsymbol{X}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{X})^{-1}$ does not exist

    $\tilde{\beta}$ is a biased estimator: $\mathbb{E}_{\boldsymbol{y}}[\tilde{\beta}\,\big|\,\boldsymbol{S}] \neq \beta_0$

    Variance $\mathbb{V}\mathrm{ar}_{\boldsymbol{y}}[\tilde{\beta}\,\big|\,\boldsymbol{S}]$ is singular

Combined Uncertainty:
Uniform Sampling with Replacement (1)

Sampling 2 out of 4 rows: $\boldsymbol{S}_{ij} = \sqrt{2} \begin{pmatrix} \boldsymbol{e}_i^T \\ \boldsymbol{e}_j^T \end{pmatrix}, \ 1 \le i, j \le 4$

$$\boldsymbol{S}_{11}\boldsymbol{X} = \sqrt{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\boldsymbol{S}_{42}\boldsymbol{X} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Unbiased estimator of the identity:

$$\mathbb{E}_{\boldsymbol{s}}[\boldsymbol{S}^T\boldsymbol{S}] = \sum_{i=1}^{4}\sum_{j=1}^{4} \tfrac{1}{16}\boldsymbol{S}_{ij}^T\boldsymbol{S}_{ij} = \boldsymbol{I} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

Combined Uncertainty: Uniform Sampling with Replacement (2)

$\min_\beta \| \boldsymbol{S}(\boldsymbol{X}\beta - \boldsymbol{y}) \|_2$ has solution $\tilde{\beta} = (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\boldsymbol{y}$

Conditional mean: $\mathbb{E}_{\boldsymbol{y}}[\tilde{\beta} \mid \boldsymbol{S}] = (\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\boldsymbol{X}\beta_0$

Sequential conditioning:

$$\mathbb{E}[\tilde{\beta}] = \mathbb{E}_{\boldsymbol{s}} \left[\mathbb{E}_{\boldsymbol{y}} \left[\tilde{\beta} \mid \boldsymbol{S} \right] \right] = \mathbb{E}_{\boldsymbol{s}} \left[(\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\boldsymbol{X} \right] \beta_0$$

Biased estimator:

$$\mathbb{E}_{\boldsymbol{s}} \left[(\boldsymbol{S}\boldsymbol{X})^\dagger \boldsymbol{S}\boldsymbol{X} \right] = \sum_{i=1}^{4} \sum_{j=1}^{4} \tfrac{1}{16} (\boldsymbol{S}_{ij}\boldsymbol{X})^\dagger (\boldsymbol{S}_{ij}\boldsymbol{X}) = \tfrac{1}{16} \begin{pmatrix} 12 & 0 \\ 0 & 7 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

12 out of 16 sketched matrices $\boldsymbol{S}_{ij}\boldsymbol{X}$ are rank deficient

Structural Perturbation Bounds

Perturbed Solution

Exact problem $\min_\beta \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2$

Hat matrix $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{X}\mathbf{X}^\dagger$

Range $\mathcal{R}(\mathbf{P}_x) = \mathcal{R}(\mathbf{X})$

Solution $\hat{\boldsymbol{\beta}} = \mathbf{X}^\dagger\mathbf{y} = \mathbf{X}^\dagger\mathbf{P}_x\mathbf{y}$

Perturbed Solution

Exact problem $\min_\beta \|\boldsymbol{X}\beta - \boldsymbol{y}\|_2$

Hat matrix $\boldsymbol{P_x} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{X}\boldsymbol{X}^\dagger$

Range $\mathcal{R}(\boldsymbol{P_x}) = \mathcal{R}(\boldsymbol{X})$

Solution $\hat{\beta} = \boldsymbol{X}^\dagger\boldsymbol{y} = \boldsymbol{X}^\dagger\boldsymbol{P_x}\boldsymbol{y}$

Perturbed problem $\min_\beta \|\boldsymbol{S}(\boldsymbol{X}\beta - \boldsymbol{y})\|_2$

$(\boldsymbol{X}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{X})^{-1}$ does not work!

Perturbed Solution

Exact problem $\min_\beta \|\boldsymbol{X}\beta - \boldsymbol{y}\|_2$

Hat matrix $\boldsymbol{P_x} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{X}\boldsymbol{X}^\dagger$

Range $\mathcal{R}(\boldsymbol{P_x}) = \mathcal{R}(\boldsymbol{X})$

Solution $\hat{\beta} = \boldsymbol{X}^\dagger\boldsymbol{y} = \boldsymbol{X}^\dagger\boldsymbol{P_x}\boldsymbol{y}$

Perturbed problem $\min_\beta \|\boldsymbol{S}(\boldsymbol{X}\beta - \boldsymbol{y})\|_2$

$(\boldsymbol{X}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{X})^{-1}$ does not work!

Comparison Hat matrix[*] $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{S}\boldsymbol{X})^\dagger\boldsymbol{S}$

Range $\mathcal{R}(\boldsymbol{P}) \subset \mathcal{R}(\boldsymbol{X}) = \mathcal{R}(\boldsymbol{P_x})$

      If $\mathrm{rank}(\boldsymbol{S}\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{X})$ then $\mathcal{R}(\boldsymbol{P}) = \mathcal{R}(\boldsymbol{X})$

Solution $\tilde{\beta} = (\boldsymbol{S}\boldsymbol{X})^\dagger\boldsymbol{S}\boldsymbol{y} = \boldsymbol{X}^\dagger\boldsymbol{P}\boldsymbol{y}$

[*]More general than [Raskutti, Mahoney 2016]

Example: Hat Matrix, and Comparison Hat Matrix

Design matrix, and Hat matrix:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{P_x} = \boldsymbol{X}\boldsymbol{X}^{\dagger} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Sketching matrix with $\mathrm{rank}(\boldsymbol{SX}) = \mathrm{rank}(\boldsymbol{X})$, and Comparison Hat matrix:

$$\boldsymbol{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{P} = \boldsymbol{X}(\boldsymbol{SX})^{\dagger}\boldsymbol{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{R}(\boldsymbol{P}) = \mathcal{R}(\boldsymbol{X})$$

Sketching matrix with $\mathrm{rank}(\boldsymbol{SX}) < \mathrm{rank}(\boldsymbol{X})$, and Comparison Hat matrix:

$$\boldsymbol{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{P} = \boldsymbol{X}(\boldsymbol{SX})^{\dagger}\boldsymbol{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{R}(\boldsymbol{P}) \subset \mathcal{R}(\boldsymbol{X})$$

Perturbed Solution

Exact problem $\min_\beta \|X\beta - y\|_2$

Hat matrix $P_x = X(X^T X)^{-1} X^T = X X^\dagger$

Solution $\hat\beta = X^\dagger P_x y$

Perturbed problem $\min_\beta \|S(X\beta - y)\|_2$

Comparison Hat matrix $P = X(SX)^\dagger S$

Solution $\tilde\beta = X^\dagger P y$

Difference between perturbed and exact solution

$$\tilde\beta = \hat\beta + X^\dagger (P - P_x) y$$

proportional to difference between Hat and Comparison Hat matrix

Multiplicative Perturbation Bounds

Ingredients:

Condition number: $\kappa_2(\boldsymbol{X}) = \|\boldsymbol{X}\|_2 \|\boldsymbol{X}^\dagger\|_2$

Angle between $\boldsymbol{y}$ and range($\boldsymbol{X}$): $0 < \theta < \pi/2$

Relative error in perturbed solution:

$$\frac{\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\|_2}{\|\hat{\boldsymbol{\beta}}\|_2} \leq \underbrace{\frac{\kappa_2(\boldsymbol{X})}{\cos\theta}}_{\text{Amplifier}} \underbrace{\|\boldsymbol{P} - \boldsymbol{P_x}\|_2}_{\text{Perturbation}}$$

Least squares solution insensitive to multiplicative perturbations, if

1. Matrix $\boldsymbol{X}$ well-conditioned with respect to (left) inversion
2. Righthand side $\boldsymbol{y}$ close to range($\boldsymbol{X}$)

Tighter than [Drineas, Mahoney, Muthukrishnan, Sarlós 2011]

Model-Induced Uncertainty, Conditioned on Algorithm-Induced uncertainty

Conditioning on S: Mean

$\min_\beta \|SX\beta - Sy\|_2$ has solution $\tilde{\beta} = (SX)^\dagger S y$

Conditional mean

$$\mathbb{E}_y\left[\tilde{\beta} \,\Big|\, S\right] = (SX)^\dagger S \, \mathbb{E}_y[y] = (SX)^\dagger S X \, \beta_0$$

- If $\mathrm{rank}(SX) = \mathrm{rank}(X)$ then $\tilde{\beta}$ is unbiased estimator

$$\mathbb{E}_y\left[\tilde{\beta} \,\Big|\, S\right] = \underbrace{(SX)^\dagger (SX)}_{I} \, \beta_0 = \beta_0$$

- If $\mathrm{rank}(SX) < \mathrm{rank}(X)$ then $\tilde{\beta}$ is biased estimator

$$\mathbb{E}_y\left[\tilde{\beta} \,\Big|\, S\right] = \beta_0 + \underbrace{\left(I - (SX)^\dagger (SX)\right)}_{\text{Rank deficiency of } SX} \beta_0$$

Bias of $\tilde{\beta}$ increases with rank deficiency of SX

Conditioning on $\boldsymbol{S}$: Variance

$\min_\beta \|\boldsymbol{S}\boldsymbol{X}\beta - \boldsymbol{S}\boldsymbol{y}\|_2$ has solution $\tilde{\beta} = \boldsymbol{X}^\dagger \boldsymbol{P}\boldsymbol{y}$

Ingredients:

   Hat matrix: $\boldsymbol{P_x} = \boldsymbol{X}\boldsymbol{X}^\dagger$
   Comparison Hat matrix: $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{S}\boldsymbol{X})^\dagger\boldsymbol{S}$
   Model variance: $\mathbb{V}\mathrm{ar}_{\boldsymbol{y}}[\hat{\beta}] = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \sigma^2\boldsymbol{X}^\dagger\left(\boldsymbol{X}^\dagger\right)^T$

Conditional variance:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}_{\boldsymbol{y}}\left[\tilde{\beta}\,\middle|\,\boldsymbol{S}\right] &= \sigma^2\,\boldsymbol{X}^\dagger\boldsymbol{P}\boldsymbol{P}^T(\boldsymbol{X}^\dagger)^T \\
&= \mathbb{V}\mathrm{ar}_{\boldsymbol{y}}[\hat{\beta}] + \underbrace{\sigma^2\,\boldsymbol{X}^\dagger\left(\boldsymbol{P}\boldsymbol{P}^T - \boldsymbol{P_x}\right)(\boldsymbol{X}^\dagger)^T}_{\text{Deviation of }\boldsymbol{P}\text{ from orthogonal projector}}
\end{aligned}
$$

$\boldsymbol{P_x}$ is orthogonal projector onto $\mathcal{R}(\boldsymbol{X})$ with ${\boldsymbol{P_x}}^T = \boldsymbol{P_x} = {\boldsymbol{P_x}}^2$

Conditioning on S: Summary

$\min_{\beta} \|SX\beta - Sy\|_2$ has solution $\tilde{\beta} = X^{\dagger}Py$
Comparison Hat matrix $P = X(SX)^{\dagger}S$

Model-induced uncertainty of $\tilde{\beta}$ conditioned on S
governed by $\mathrm{rank}(SX)$

- Bias increases with deviation of SX from full column-rank
- If $\mathrm{rank}(SX) = \mathrm{rank}(X)$ then $\tilde{\beta}$ is unbiased
- Conditional variance close to model variance, if
  P close to being an orthogonal projector onto $\mathcal{R}(X)$
  X well-conditioned with respect to inversion

- If $\mathrm{rank}(SX) < \mathrm{rank}(X)$ then $\mathbb{V}\mathrm{ar}_y[\tilde{\beta} \mid S]$ is singular

Combined Model-induced and Algorithm-Induced Uncertainty

Combined Uncertainty: Mean

$\min_\beta \|\boldsymbol{SX}\beta - \boldsymbol{Sy}\|_2$ has solution $\tilde{\beta} = (\boldsymbol{SX})^\dagger \boldsymbol{S\,y}$

Total mean

$$\mathbb{E}[\tilde{\beta}] = \beta_0 + \mathbb{E}_s \left[(\boldsymbol{SX})^\dagger (\boldsymbol{SX}) - \boldsymbol{I} \right] \beta_0$$

- Deviation of combined uncertainties from model-induced uncertainties governed by
  expected deviation of sketched matrix from rank deficiency

- Total bias proportional to
  expected deviation of $\boldsymbol{SX}$ from having full column-rank

Combined Uncertainty: Variance

$\min_{\beta} \| \boldsymbol{SX}\beta - \boldsymbol{Sy} \|_2$ has solution $\tilde{\beta} = (\boldsymbol{SX})^{\dagger} \boldsymbol{S} \, \boldsymbol{y}$

Deviation of total variance from model variance

$$\mathbb{Var}[\tilde{\beta}] = \mathbb{Var}_{\boldsymbol{y}}[\hat{\beta}] + \sigma^2 \, \boldsymbol{X}^{\dagger} \, \mathbb{E}_{\boldsymbol{s}}[\boldsymbol{PP}^T - \boldsymbol{P_x}] (\boldsymbol{X}^{\dagger})^T$$
$$+ \mathbb{Var}_{\boldsymbol{s}} \left[\left((\boldsymbol{SX})^{\dagger}(\boldsymbol{SX}) - \boldsymbol{I} \right) \beta_0 \right]$$

- $\boldsymbol{X}^{\dagger} \, \mathbb{E}_{\boldsymbol{s}}[\boldsymbol{PP}^T - \boldsymbol{P_x}] (\boldsymbol{X}^{\dagger})^T$
  Expected deviation of $\boldsymbol{P}$ from orthogonal projector onto $\mathcal{R}(\boldsymbol{X})$,
  amplified by conditioning of $\boldsymbol{X}$

- $\mathbb{Var}_{\boldsymbol{s}} \left[((\boldsymbol{SX})^{\dagger}(\boldsymbol{SX}) - \boldsymbol{I}) \beta_0 \right]$
  Expected deviation of $\boldsymbol{SX}$ from having full column-rank

Example: Best Case for Uniform Sampling

Columns of Hadamard matrix: Best Coherence

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad \boldsymbol{P_x} = \boldsymbol{X}\boldsymbol{X}^{\dagger} = \tfrac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$\boldsymbol{S}$ samples 2 rows uniformly and with replacement

- Expected deviation of $\boldsymbol{SX}$ from full column-rank

$$\mathbb{E}_{\boldsymbol{s}}\left[(\boldsymbol{SX})^{\dagger}(\boldsymbol{SX}) - \boldsymbol{I}\right] = -\tfrac{4}{16}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- Expected deviation of $\boldsymbol{P}$ from orthogonal projector onto $\mathcal{R}(\boldsymbol{X})$

$$\mathbb{E}_{\boldsymbol{s}}[\boldsymbol{PP}^{T} - \boldsymbol{P_x}] = \tfrac{3}{16}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Example: Worst Case for Uniform Sampling

Columns of identity matrix: Worst Coherence

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{P_x} = \boldsymbol{X}\boldsymbol{X}^\dagger = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$\boldsymbol{S}$ samples 2 rows uniformly and with replacement

- Expected deviation of $\boldsymbol{SX}$ from full column-rank

$$\mathbb{E}_{\boldsymbol{s}}\left[(\boldsymbol{SX})^\dagger(\boldsymbol{SX}) - \boldsymbol{I}\right] = -\tfrac{9}{16}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- Expected deviation of $\boldsymbol{P}$ from orthogonal projector onto $\mathcal{R}(\boldsymbol{X})$

$$\mathbb{E}_{\boldsymbol{s}}[\boldsymbol{PP}^T - \boldsymbol{P_x}] = \tfrac{9}{16}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Summary

- Randomized (row) sketching for full column-rank regression

- Exact expressions for uncertainties, induced by model and algorithm, under very general assumptions

- Introduced Comparison Hat matrix, to allow comparison between problems of different dimensions

- Tighter multiplicative perturbation bounds

- Total mean and variance governed by expected deviation of

    Sketched matrix from full column-rank
    Comparison Hat matrix from orthogonal projector

- Examples illustrate applicability

- J.T. Chi and I.C.F. Ipsen,
  Randomized Least Squares Regression: Combining Model- and Algorithm-Induced Uncertainties, arXiv:1808.0594

MathJobs.Org

Department of Mathematics, North Carolina State University

Position ID: NCSU-PDRNA [#12087]
Position Title: RTG in Randomized Numerical Analysis, Postdoctoral Fellow
Position Type: Postdoctoral
Position Location: Raleigh, North Carolina 27695, United States [map]
Subject Area: Randomized numerical analysis
Application Deadline: 2019/03/15 (posted 2018/06/27)
Position Description: Apply

The Department of Mathematics at NC State University has an open postdoctoral position under its new NSF funded Research Training Group RTG, DMS-1745654 in Randomized Numerical Analysis. The position is restricted to US citizens, nationals and permanent residents. The postdoctoral associate must hold a PhD in mathematics or a closely related discipline awarded after January 1, 2017. An exception to this rule, occasioned for instance by a career interruption, may be requested in the cover letter.

The postdoctoral associate is expected to engage in research related to at least one of the three complementary research thrusts of the project in randomized numerical linear algebra, random nonlinear solvers and global sensitivity analysis. Strengths in applied probability, numerical analysis and scientific computing are an asset. Based on background, experience and interests, the postdoc will be assigned one research mentor from the faculty associated with the RTG: Alen Alexanderian, Pierre Gremaud, Ilse Ipsen, Tim Kelley, Arvind Saibaba and Ralph Smith. An academic mentor will be chosen from this group to oversee the postdoc's training and professional development. The postdoc will also play a leading role in mentoring students associated with some of the project's working groups; he/she will be expected to assist in organizing seminars and activities related to the general research of the RTG. In addition, the postdoc will teach on average one course per semester in the Department of Mathematics.

This position is for a term of up to three years, contingent upon performance. The starting date is anticipated between January and August 2019. The 11 month salary of $62,500 includes summer support. In addition, the postdoc will be given an annual travel budget of $3,000.

Through the mathjobs.org system, applicants should submit a cover letter, a CV, a research statement and the names of at least three potential letter writers. Women and members of under-represented minority groups are particularly encouraged to apply. Applications will be evaluated on a rolling basis until the position is filled; please apply as soon as possible. Inquiries may be directed to Pierre Gremaud (RTG Director): gremaud@ncsu.edu.

Data Science

Editor-in-Chief
Ilse Ipsen, North Carolina State University

Editorial Board
Amr El-Bakry, ExxonMobil
Nicholas J. Higham, University of Manchester
Michael Mahoney, UC Berkeley

Data Science, SIAM's newest book series, covers the mathematical, computational, and scientific aspects of data science. The series will publish high-impact research monographs, in-depth essays on emerging trends, tutorials with a broad reach, state-of-the-art surveys, scholarly research retrospectives, and textbooks.