A Hybrid ADMM for Big Data Applications

Daniel P. Robinson and Rachael Tappenden
Johns Hopkins University
Department of Applied Mathematics and Statistics

Funding: National Science Foundation

University of Edinburgh
Optimization and Big Data
May 7, 2015

Gauss-Seidel versus Jacobi ADMM

The problem:

$$\underset{x_1,\ldots,x_n}{\text{minimize}} \;\; f(x) := \sum_{i=1}^{n} f_i(x_i) \;\; \text{subject to} \;\; \sum_{i=1}^{n} A_i x_i = b$$

Gauss-Seidel ADMM (GS-ADMM)

Solve n problems during the kth iteration of the form

$$x_j^{(k+1)} = \underset{x_j}{\text{argmin}} \;\; \mathcal{L}(x_1^{(k+1)}, \ldots x_{j-1}^{(k+1)}, x_j, x_{j+1}^{(k)}, \ldots, x_n^{(k)}; y^{(k)}) + \tfrac{1}{2}\|x_j - x_j^{(k)}\|_{P_j}^2$$

- serial in nature
- use up-to-date information
- time $\approx n$ minimizations

Jacobi ADMM (J-ADMM) (Deng, Lai, Peng, and Yin)

Solve n problems during the kth iteration of the form

$$x_j^{(k+1)} = \underset{x_j}{\text{argmin}} \;\; \mathcal{L}(x_1^{(k)}, \ldots x_{j-1}^{(k)}, x_j, x_{j+1}^{(k)}, \ldots, x_n^{(k)}; y^{(k)}) + \tfrac{1}{2}\|x_j - x_j^{(k)}\|_{P_j}^2$$

- can be done in parallel
- use old information
- time ≈ 1 minimization … if you have $\geq n$ machines!

Hybrid ADMM

The problem:

$$\underset{x_1,\ldots,x_n}{\text{minimize}}\ f(x) := \sum_{i=1}^{n} f_i(x_i) \quad \text{subject to} \quad \sum_{i=1}^{n} A_i x_i = b$$

Group the data: Form ℓ groups of p blocks (number of machines):

$$x \quad = \quad [\underbrace{x_1,\ldots,x_p}_{\mathbf{x}_1} \mid \underbrace{x_{p+1},\ldots,x_{2p}}_{\mathbf{x}_2} \mid \ \ldots \ \mid \underbrace{x_{lp+1}\ldots x_n}_{\mathbf{x}_l}]$$

$$A \quad = \quad [\underbrace{A_1,\ldots,A_p}_{\mathcal{A}_1} \mid \underbrace{A_{p+1},\ldots,A_{2p}}_{\mathcal{A}_2} \mid \ \ldots \ \mid \underbrace{A_{lp+1}\ldots A_n}_{\mathcal{A}_l}]$$

Group regularization:

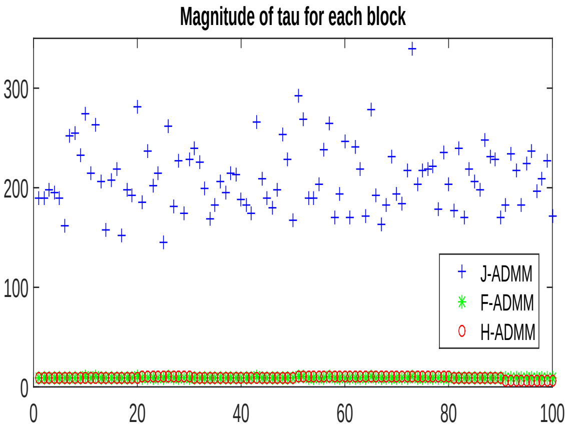
$$\mathcal{P}_i := \text{diag}(P_{(i-1)p+1},\ldots,P_{ip}) - \rho \mathcal{A}_i^T \mathcal{A}_i \quad \text{for } i = 1,\ldots,\ell$$

makes group iterations separable/parallelizable!

- Jacobi updating within a group
- Gauss-Seidel updating between groups
- Cost $\approx \ell$ minimizations (same as full Jacobi updating!)

Results (theoretical values for regularization matrices)

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; \frac{1}{2}\|x\|_2^2 \quad \text{subject to} \quad Ax = b$$



Method	Epochs
J-ADMM	4358.2
H-ADMM	214.1
F-ADMM	211.3

Results (tuning the regularization matrices)

τ_i	J-ADMM	H-ADMM	F-ADMM
$\frac{\rho^2}{2}\|A\|^4$	530.0	526.3	526.2
$0.6 \cdot \frac{\rho^2}{2}\|A\|^4$	324.0	320.1	319.9
$0.4 \cdot \frac{\rho^2}{2}\|A\|^4$	217.7	214.5	214.1
$0.22 \cdot \frac{\rho^2}{2}\|A\|^4$	123.1	119.3	119.0
$0.2 \cdot \frac{\rho^2}{2}\|A\|^4$	—	95.8	95.5
$0.1 \cdot \frac{\rho^2}{2}\|A\|^4$	—	75.3	73.0

Bottom line:

- H-ADMM uses a combination of Gauss-Seidel and Jacobi updating in order to "optimize" computation
- allows most recent updates to be used between groups
- H-ADMM has the same cost as J-ADMM
- numerical results show the practical advantage of using updated information