

IMPROVING THE RANDOM BLOCK-COORDINATE FORWARD-BACKWARD METHOD

$$\min_{x \in \mathcal{H}} f(x) + g(x)$$

$$\mathcal{H} = \bigoplus_{i=1}^m \mathcal{H}_i \quad x = (x_1, \dots, x_m) \quad [\text{Ex. } \mathbb{R}^m = \mathbb{R} \times \dots \times \mathbb{R}]$$

$$g: \mathcal{H} \rightarrow]-\infty, +\infty] \quad g(x) = \sum_{i=1}^m g_i(x_i), \quad g_i \in \Gamma_0(\mathcal{H}_i)$$

$$f: \mathcal{H} \rightarrow \mathbb{R} \text{ convex differentiable}$$

Example (Lasso problem)

$$\min_{x \in \mathbb{R}^m} \frac{1}{2} \|Ax - b\|^2 + \lambda \sum_{i=1}^m |x_i|$$

Algorithm

FBA

$$x^{k+1} = \text{prox}_{\gamma g}(x^k - \gamma \nabla f(x^k))$$

$$\begin{cases} \forall i \in \{1, \dots, m\} \\ x_i^{k+1} = \text{prox}_{\gamma g_i}(x^k - \gamma \nabla_i f(x^k)) \end{cases}$$

$$\nabla f: \mathcal{H} \rightarrow \mathcal{H}, \quad \nabla f(x) = (\nabla_1 f(x), \dots, \nabla_m f(x))$$

RBC-FBA

$$\begin{aligned} \forall i \in \{1, \dots, m\} \quad x_i^{k+1} &= x_i^k + \varepsilon_i^k [\text{prox}_{\gamma_i g_i}(x_i^k - \gamma_i \nabla_i f(x^k))] \\ &= \begin{cases} \text{prox}_{\gamma_i g_i}(x_i^k - \gamma_i \nabla_i f(x^k)) & \text{if } \varepsilon_i^k = 1 \\ x_i^k & \text{if } \varepsilon_i^k = 0 \end{cases} \end{aligned}$$

where ε_i^k are Bernoulli random variables with values in $\{0, 1\}$
 $\varepsilon^k = (\varepsilon_1^k, \dots, \varepsilon_m^k) \in \{0, 1\}^m$ and $(\varepsilon^k)_{k \in \mathbb{N}}$ are iid

Special case $g \equiv 0$

$$\forall i \in \{1, \dots, m\} \quad x_i^{k+1} = x_i^k - \gamma_i \varepsilon_i^k \nabla_i f(x^k).$$

Remark The stepsize is allowed to depend on the blocks

- stepsize rule for FBA : $\gamma < \frac{2}{L}$, L is the Lipschitz constant of ∇f .
- stepsize rule for RBC-FBA : $\gamma_i < \frac{2}{\tau L_i}$, where
- the L_i 's are the block-coordinate Lipschitz constants of ∇f , meaning

$$\forall i \quad \|\nabla_i f(x_1, \dots, x_i', \dots, x_m) - \nabla_i f(x_1, \dots, x_i, \dots, x_m)\| \leq L_i \|x_i' - x_i\|$$

The L_i 's may be much smaller than L .

- τ is the number of blocks updated in parallel : $P(\sum_{i=1}^m \varepsilon_i^k = \tau) = 1$

Example (Lasso problem) $f(x)$

$$\min_{x \in \mathbb{R}^m} \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 \quad \mathcal{H} = \mathbb{R}^m$$

$$A = [a^1 \dots a^m] \in \mathbb{R}^{p \times m}, \quad b \in \mathbb{R}^p$$

$$\nabla f(x) = A^T (Ax - b) = \begin{bmatrix} a^{1T}(Ax - b) \\ \vdots \\ a^{mT}(Ax - b) \end{bmatrix}$$

$$\underline{\underline{L_i = \|a^i\|^2}}$$

RBC-FBA

$$x^{k+1} = x^k + \sum_{i=1}^m \varepsilon_i^k \underbrace{\left[\text{soft}_{\gamma_i \lambda} (x_i^k - \gamma_i a^{iT}(Ax^k - b)) - x_i^k \right]}_{\in \mathbb{R}} e_i$$

Remark

The stepsize rule for RBC-FBA is $\gamma_i < \frac{2}{\varepsilon \|a^i\|^2}$

The stepsize rule for FBA is $\gamma_i < \frac{2}{\|A\|^2}$

In the worst case $\|A\| = m \|a^i\|$, but it may be $\|A\| \gg \max_{1 \leq i \leq m} \|a^i\|$

Convergence results $F = f + g$. $\|x\|_W^2 = \sum_{i=1}^m \frac{1}{r_i p_i} \|x_i\|^2$. $p_i = P(\varepsilon_i^k = 1)$. $r_i = \frac{1}{\beta L_i}$

1. $E[F(x^k)] - \inf F = o(1/k)$

2. $x^k \rightarrow x_* \in \operatorname{argmin} F$ P-a.s.

3. $E[F(x^k)] - \inf F \leq \left[\frac{\|x^0 - x_*\|_W^2}{2} + \left(\frac{1}{p_{\min}} - 1 \right) (E[F(x^0)] - \inf F) \right] \frac{1}{k}$

3a. If $p_i = p = \frac{\gamma}{m}$

$$E[F(x^{km/\gamma})] - \inf F \leq \left[\frac{\gamma}{2} \sum_{i=1}^m L_i \|x_i^0 - x_{*i}\|^2 + \left(1 - \frac{\gamma}{m} \right) (E[F(x^0)] - \inf F) \right] \frac{1}{k}$$

$$[FBA : F(x^k) - \inf F \leq \frac{L}{2} \|x^0 - x_*\|^2]$$

3b. If $p_i = \gamma L_i / \sum_{j=1}^m L_j$ and $\bar{L} = \sum_{j=1}^m L_j / m$,

$$E[F(x^{km/\gamma})] - \inf F \leq \left[\min\{\gamma, \gamma\} \bar{L} \frac{\|x^0 - x_*\|^2}{2} + \left(\frac{\bar{L}}{L_{\min}} - \gamma \right) (E[F(x^0)] - \inf F) \right] \frac{1}{k}$$

4. If $\operatorname{dist}_{\tilde{r}^{-1}}(x, \operatorname{argmin} F) \leq c \|x - \operatorname{prox}_g^{\tilde{r}^{-1}}(x - \nabla^{\tilde{r}^{-1}} f(x))\|_{\tilde{r}^{-1}}$ Luo-Tseng EB

$$E[F(x^k)] - \inf F \leq \left(1 - \frac{p_{\min}}{2c} \right)^k (E[F(x^0)] - \inf F)$$

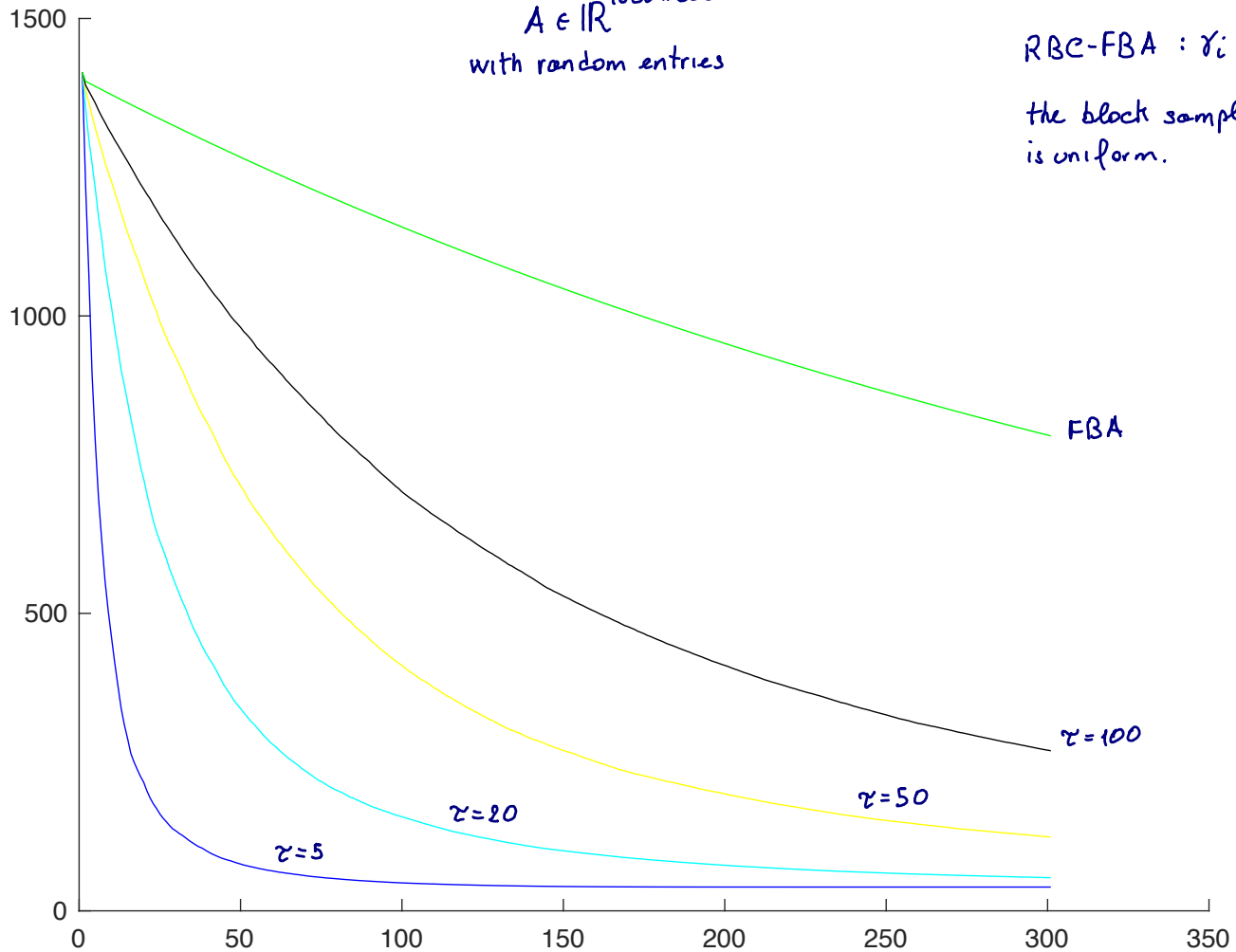
$$x^k \rightarrow x_* \in \operatorname{argmin} F \text{ and } E[\|x^k - x_*\|_W] \leq \left(1 - \frac{p_{\min}}{2c} \right)^{k/2} \text{const.}$$

LASSO PROBLEM
 $A \in \mathbb{R}^{1000 \times 500}$
 with random entries

FBA : $\gamma = \frac{1}{\|A\|^2}$

RBC-FBA : $\gamma_i = \frac{1}{2\|a_i\|^2}$

the block sampling
 is uniform.



Example

$$(P) \quad \min_{Ax=b} \frac{1}{2} \|x\|^2$$

$$A \in \mathbb{R}^{p \times m} \quad b \in \mathcal{R}(A) \subset \mathbb{R}^p$$

$$(D) \quad \min_{u \in \mathbb{R}^m} \frac{1}{2} \|A^T u\|^2 - u^T b := D(u) \quad A^T = [a_1 \dots a_m] \quad L_i = \|a_i\|^2$$

$$x = A^T u : \quad \frac{1}{2} \|x - x_*\|^2 \leq D(u) - \inf D$$

$$\text{RBC-FBA} : \quad u^{k+1} = u^k - \sum_{i=1}^m \varepsilon_i^k \gamma_i (a_i^T A^T u^k - b_i) a_i \quad \gamma_i < \frac{2}{\sum \|a_i\|^2}$$

multiplying by A^T :
$$x^{k+1} = x^k - \sum_{i=1}^m \varepsilon_i^k \gamma_i (a_i^T A^T u^k - b_i) a_i$$
 $\leftarrow$ SGD on $\min_x \sum_{i=1}^m (a_i^T x - b_i)^2$

from 4.
$$\frac{1}{2} E[\|x^k - x_*\|^2] \leq \left(1 - \underbrace{P_{\min} \gamma_{\min}}_{\text{extra factor!}} \underbrace{\sigma_{\min}^2(A)}_{\text{they should be coupled!}}\right)^k \text{const.}$$

Special case $\gamma = 1$
$$\gamma_i = \frac{1}{\|a_i\|^2}$$

$$x^{k+1} = x^k - \frac{a_{i_k}^T x^k - b_{i_k}}{\|a_{i_k}\|^2} \cdot a_{i_k}$$

Randomized Kaczmarz method
by Strohmer and Vershynin

$$P_i = \frac{\|a_i\|^2}{\|A\|_F^2}$$

$$\frac{1}{2} E[\|x^k - x_*\|^2] \leq \left(1 - \underbrace{\frac{1}{2} \frac{\sigma_{\min}^2(A)}{\|A\|_F^2} \cdot \frac{\min \|a_i\|^2}{\max \|a_i\|^2}}_{\text{not tight!}}\right)^k \text{const}$$

$$\left[\left(1 - \frac{\sigma_{\min}^2(A)}{\|A\|_F^2}\right)^k \right]$$

CONVERGENCE ANALYSIS

$$\bar{x}^{k+1} = (\text{prox}_{\gamma_i g_i}(x^k - \gamma_i \nabla_i f(x^k)))_{1 \leq i \leq m} = \text{prox}_g^{\bar{\Gamma}^1}(x^k - \nabla^{\bar{\Gamma}^1} f(x^k)), \quad \|x\|_{\bar{\Gamma}^1}^2 = \sum_{i=1}^m \frac{1}{\gamma_i} \|x_i\|^2$$

LEMMA -
$$\| \bar{x}^{k+1} - x^k \|_{\bar{\Gamma}^1}^2 \leq \|x^k - x^k\|_{\bar{\Gamma}^1}^2 - \| \bar{x}^{k+1} - x^k \|_{\bar{\Gamma}^1}^2 + 2 \underbrace{(g(x^k) - g(\bar{x}^{k+1}) + \langle \nabla f(x^k), x^k - \bar{x}^{k+1} \rangle)}_{(a) \geq 0} - 2(F(x^k) - F(x))$$

DESCENT LEMMA {
$$\mathbb{E}[f(x^{k+1}) - f(x^k) | \varepsilon^0, \dots, \varepsilon^{k-1}] \leq \mathbb{E}[\langle \nabla f(x^k), x^{k+1} - x^k \rangle | \varepsilon^0, \dots, \varepsilon^{k-1}] + \frac{1}{2} \rho_i \sum L_i \|x_i^k - \bar{x}_i^{k+1}\|^2$$

The problem is majorizing
$$g(x^k) - g(\bar{x}^{k+1}) + \langle \nabla f(x^k), x^k - \bar{x}^{k+1} \rangle$$

$$g(x^k) - g(\bar{x}^{k+1}) + \langle \nabla f(x^k), x^k - \bar{x}^{k+1} \rangle = \mathbb{E} \left[\sum_{i=1}^m \frac{\varepsilon_i^k}{\rho_i} (g_i(x_i^k) - g_i(\bar{x}_i^{k+1}) + \langle \nabla_i f(x^k), x_i^k - \bar{x}_i^{k+1} \rangle) | \varepsilon^0, \dots, \varepsilon^{k-1} \right]$$

and
$$\sum_{i=1}^m \frac{\varepsilon_i^k}{\rho_i} (g_i(x_i^k) - g_i(\bar{x}_i^{k+1}) + \langle \nabla_i f(x^k), x_i^k - \bar{x}_i^{k+1} \rangle)$$

$\leq \dots$

$$\leq \frac{1}{\rho_{\min}} (g(x^k) - g(\bar{x}^{k+1}) + \langle \nabla f(x^k), x^k - \bar{x}^{k+1} \rangle) - \sum_{i=1}^m \underbrace{\left(\frac{1}{\rho_{\min}} - \frac{1}{\rho_i} \right)}_{\geq 0} \frac{\varepsilon_i^k}{\gamma_i} \|x_i^k - \bar{x}_i^{k+1}\|^2$$

$$\underbrace{g(x^k) - g(\bar{x}^{k+1}) + \langle \nabla f(x^k), x^k - \bar{x}^{k+1} \rangle}_{(a)} \leq \frac{1}{\rho_{\min}} \mathbb{E}[F(x^k) - F(x^{k+1}) | \varepsilon^0, \dots, \varepsilon^{k-1}] + \frac{1}{2} \| \bar{x}^{k+1} - x^k \|_{\bar{\Gamma}^1}^2$$

CONVERGENCE ANALYSIS II (for error bounds)

FUNDAMENTAL
INEQUALITY

$$\| \bar{x}^{k+1} - x \|_{\tilde{r}^{-1}}^2 \leq \| x^k - x \|_{\tilde{r}^{-1}}^2 + \frac{2}{\rho_{\min}} \mathbb{E} [F(x^k) - F(x^{k+1}) \mid \varepsilon^0, \dots, \varepsilon^{k-1}] - 2(F(x^k) - F(x))$$

$$\Rightarrow \| \bar{x}^{k+1} - x \|_{\tilde{r}^{-1}}^2 \leq \frac{2}{\rho_{\min}} \mathbb{E} [F(x^k) - F(x^{k+1}) \mid \varepsilon^0, \dots, \varepsilon^{k-1}]$$

$$1. \| x^k - x \|_{\tilde{r}^{-1}}^2 - \| \bar{x}^{k+1} - x \|_{\tilde{r}^{-1}}^2 = - \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}}^2 + 2 \langle x^k - \bar{x}^{k+1}, x^k - x \rangle_{\tilde{r}^{-1}}$$

$$2. \frac{2}{\rho_{\min}} \mathbb{E} [F(x^{k+1}) - F(x^k) \mid \varepsilon^0, \dots, \varepsilon^{k-1}]$$

$$\leq - \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}}^2 + 2 \langle x^k - \bar{x}^{k+1}, x^k - x \rangle_{\tilde{r}^{-1}} - 2(F(x^k) - F(x))$$

$$\leq - \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}}^2 + 2 \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}} \| x^k - x \|_{\tilde{r}^{-1}} - 2(F(x^k) - F(x))$$

$$= - \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}}^2 + 2 \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}} \text{dist}_{\tilde{r}^{-1}}(x^k, \arg\min F) - 2(F(x^k) - F(x))$$

$$\leq (2e-1) \| x^k - \bar{x}^{k+1} \|_{\tilde{r}^{-1}}^2 - 2(F(x^k) - F(x))$$

$$\leq \frac{2(2e-1)_+}{\rho_{\min}} \mathbb{E} [F(x^k) - F(x^{k+1}) \mid \varepsilon^0, \dots, \varepsilon^{k-1}] - 2(F(x^k) - F(x))$$

$$3. \underline{(1 + (2e-1)_+)} \mathbb{E} [F(x^{k+1}) - F_* \mid \varepsilon^0, \dots, \varepsilon^{k-1}] \leq (1 + (2e-1)_+) (F(x^k) - F_*) - \rho_{\min} (F(x^k) - F_*)$$

recursive inequality

The difficulties are mainly related to:

1. the presence of g .
 2. the parallel setting ($\tau > 1$)
-

Thank you very much!