UNIVERSITY OF MINNESOTA
Driven to Discover℠

Online Alternating Direction Method

Huahua Wang, Arindam Banerjee, ``Online alternating direction method,'' arXiv: 1306.3721v2 [cs.LG], July 2013.

Presenter: Morteza Mardani

February 24, 2014

Motivation

- Streaming ``Big Data'' analytics → Online optimization
- Formulation of interest

$$\min_{\mathbf{x},\mathbf{z}} \quad \sum_{t=1}^{T} (f_t(\mathbf{x}) + g(\mathbf{z})) \quad \text{s.to} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$$

$$\mathbf{x} \in \mathbb{R}^{n_1}, \quad \mathbf{z} \in \mathbb{R}^{n_2}$$

$$\mathbf{A} \in \mathbb{R}^{m \times n_1}, \quad \mathbf{B} \in \mathbb{R}^{m \times n_2}$$

$f_t(\mathbf{x}),\ g(\mathbf{z})$: closed proper convex

- General enough, other constraints can be included …
- Applications: distributed optimization, machine learning (e.g., LASSO),…
- Challenge: costly projections per iteration → double-loop algorithm!?

Road ahead

- **Batch ADM: convergence rate**
- **Online ADM (OADM): regret analysis**
  - **Bregman divergence: convex and strongly convex**
  - **No Bregman divergence: convex and strongly convex**
- **Inexact OADM**
- **Stochastic OADM**

Batch ADM

- $f_t = f, \ \forall t$ → $\min_{\mathbf{x},\mathbf{z}} \quad f(\mathbf{x}) + g(\mathbf{z}) \ \ \text{s.to} \ \ \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$
- Augmented Lagrangian ($\rho > 0$)

$$L_\rho(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|^2$$

- Algorithm per iteration *t* (given $\{\mathbf{z}_t, \mathbf{y}_t\}$)

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ f(\mathbf{x}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2 \},$$

$$\mathbf{z}_{t+1} = \arg\min_{\mathbf{z}} \{ g(\mathbf{z}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c}\|^2 \},$$

$$\mathbf{y}_{t+1} = \mathbf{y}_t + \rho(\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c})$$

- Equality constraint not satisfied per iteration ?!

Q: How many iterations *k* needed to obtain a *ϵ*-optimal solution?

Convergence rate

- (as1) (a) Optimal $\{\mathbf{x}^*, \mathbf{z}^*, \mathbf{y}^*\}$ exists, $\|\mathbf{y}^*\|_2 = D_y, \ \|\mathbf{z}^*\|_2 = D_z$,
  (b) $\mathbf{z}_0 = \mathbf{0}, \ \mathbf{y}_0 = \mathbf{0}$
- No smoothness assumption (needed for online scenario)

$$\bar{\mathbf{x}}_T := \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \qquad \bar{\mathbf{z}}_T := \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t$$

Theorem 1: For the iterates $\{\mathbf{x}_t, \mathbf{z}_t, \mathbf{y}_t\}$, and any feasible $\{\mathbf{x}^*, \mathbf{z}^*\}$

$$\left[f(\bar{\mathbf{x}}_T) + g(\bar{\mathbf{z}}_T)\right] - \left[f(\mathbf{x}^*) + g(\mathbf{z}^*)\right] \leq \frac{\lambda_{\max}^B D_z^2 \rho}{2T}$$

$$\lambda_{\max}^B = \lambda_{\max}(\mathbf{B}\mathbf{B}^\top)$$

- Not enough since $\{\bar{\mathbf{x}}_T, \bar{\mathbf{z}}_T\}$ may not be feasible

Proof of Theorem 1

Lemma 1: For the iterates $\{\mathbf{x}_t, \mathbf{z}_t, \mathbf{y}_t\}$,

$$\big[f(\mathbf{x}_{t+1}) + g(\mathbf{z}_{t+1})\big] - \big[f(\mathbf{x}^*) + g(\mathbf{z}^*)\big] \le \frac{1}{2\rho}\left(\|\mathbf{y}_t\|^2 - \|\mathbf{y}_{t+1}\|^2\right)$$
$$-\frac{\rho}{2}\|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2 + \frac{\rho}{2}\left(\|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_t)\|^2 - \|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_{t+1})\|\right)$$

Proof: Using the subgraident inequality

$$\in \partial f(\mathbf{x}_{t+1})$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \le -\left\langle \mathbf{A}^\top(\mathbf{y}_{t+1} + \rho(\mathbf{B}\mathbf{z}_t - \mathbf{B}\mathbf{z}_{t+1})), \mathbf{x}_{t+1} - \mathbf{x}^*\right\rangle$$
$$= -\langle \mathbf{y}_{t+1}, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}^* - \mathbf{c}\rangle + \rho\langle \mathbf{B}\mathbf{z}_{t+1} - \mathbf{B}\mathbf{z}_t, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}^* - \mathbf{c}\rangle$$
$$= -\langle \mathbf{y}_{t+1}, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}^* - \mathbf{c}\rangle + \frac{\rho}{2}\big(\|\mathbf{B}\mathbf{z}^* - \mathbf{B}\mathbf{z}_t\|^2 - \|\mathbf{B}\mathbf{z}^* - \mathbf{B}\mathbf{z}_{t+1}\|^2$$
$$+ \|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c}\|^2 - \|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2\big)$$

Similarly for g,

$$\in \partial g(\mathbf{z}_{t+1})$$

$$g(\mathbf{z}_{t+1}) - g(\mathbf{z}^*) \le -\langle \mathbf{B}^\top \mathbf{y}_{t+1}, \mathbf{z}_{t+1} - \mathbf{z}^*\rangle = -\langle \mathbf{y}_{t+1}, \mathbf{B}(\mathbf{z}_{t+1} - \mathbf{z}^*)\rangle$$

Cont'd proof

- Adding up both sides in Lemma 1 for t=1,…,T

$$\sum_{t=0}^{T-1} \left[f(\mathbf{x}_{t+1}) + g(\mathbf{z}_{t+1}) - (f(\mathbf{x}^*) + g(\mathbf{z}^*)) \right]$$

$$\leq \frac{1}{2\rho}(\|\mathbf{y}_0\|^2 - \|\mathbf{y}_T\|^2) + \frac{\rho}{2}\left(\|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_0)\|^2 - \|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_T)\|^2\right)$$

$$\leq \frac{\rho}{2}\lambda_{\max}^B D_z^2$$

- The rest follows readily from convexity …

Constraint violation

- $\{\mathbf{x}_{t+1}, \mathbf{y}_{t+1}, \mathbf{z}_{t+1}\}$ is an optimal sol. if

$$\mathbf{B}(\mathbf{z}_{t+1} - \mathbf{z}_t) = \mathbf{0}$$
$$\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c} = \mathbf{0}$$

- Residual function

$$R(s,t) := \|\mathbf{A}\mathbf{x}_s + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2 + \|\mathbf{B}\mathbf{z}_t - \mathbf{B}\mathbf{z}_{s-1}\|^2, \quad s \in \{t, t+1\}$$

Theorem 2: For the iterates $\{\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t\}$

$$R(T,T) \leq R(T,T-1) \leq \frac{\lambda_{\max}^B D_z^2 + D_y^2/\rho^2}{T}$$

- Monotonically non-increasing residuals
- $\mathcal{O}(1/T)$ for the variational form of optimality

Online ADM (OADM)

- Formulation $\min_{\mathbf{x},\mathbf{z}} \sum_{t=1}^{T} (f_t(\mathbf{x}) + g(\mathbf{z}))$ s.to $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$

- Naïve approach (e.g., COMID)

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{z}=\mathbf{c}} \quad \{f_t(\mathbf{x}) + g(\mathbf{z}) + \eta B_\phi(\mathbf{x}, \mathbf{x}_t)\}$$

- Complex double-loop algorithm

- Augmented Lagrangian at time *t*

$$L_\rho^t(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f_t(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|^2 + \eta B_\phi(\mathbf{x}, \mathbf{x}_t)$$

OADM algorithm

$\eta \geq 0$

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ f_t(\mathbf{x}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \rangle + \frac{\rho}{2} \| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \|^2 + \eta B_\phi(\mathbf{x}, \mathbf{x}_t) \},$$

$$\mathbf{z}_{t+1} = \arg\min_{\mathbf{z}} \{ g(\mathbf{z}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2} \| \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c} \|^2 \},$$

$$\mathbf{y}_{t+1} = \mathbf{y}_t + \rho(\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c})$$

- OADM is COMID with a single ADM iteration
- Message

Regret bounds	$\eta > 0$		$\eta = 0$	
	R_1	R^c	R_2	R^c
general convex	$O(\sqrt{T})$	$O(\sqrt{T})$	$O(\sqrt{T})$	$O(\sqrt{T})$
strongly convex	$O(\log T)$	$O(\log T)$	$O(\log T)$	$O(\log T)$

Regret analysis

- Algorithm presents $(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$ → nature reveals loss f_t & cons. violation
- Objective's regret

$$R_1(T) := \sum_{t=1}^{T} (f_t(\mathbf{x}_t) + g(\mathbf{z}_t)) - \min_{\mathbf{Ax}+\mathbf{Bz}=\mathbf{c}} \sum_{t=1}^{T} (f_t(\mathbf{x}) + g(\mathbf{z}))$$

- Optimality's regret

$$R^C(T) := \sum_{t=1}^{T} \left(\|\mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2 + \|\mathbf{B}(\mathbf{z}_t - \mathbf{z}_{t-1})\|^2 \right)$$

- Assumptions

(a2) Bounded subgradient, i.e., $\|f_t'(\mathbf{x})\|_q \leq G_f, \ \forall \mathbf{x} \in \mathcal{X}$

$\frac{1}{p} + \frac{1}{q} = 1$

(a3) α-strongly convex $B_\phi(\mathbf{u}, \mathbf{v}) \geq \frac{\alpha}{2}\|\mathbf{u} - \mathbf{v}\|_p^2$ for some $\alpha > 0$

(a4) $f_t(\mathbf{x}_{t+1}) + g(\mathbf{z}_{t+1}) - [f_t(\mathbf{x}^*) + g(\mathbf{z}^*)] \geq -F, \ \ \forall t$ $\quad F > 0$

Convex f_t, g ($\eta > 0$)

Theorem 2: If (a1) - (a4) hold, then for $\eta = \frac{G_f \sqrt{T}}{D_x \sqrt{2\alpha}}$ and $\rho = \sqrt{T}$,

$$R_1(T) \leq \lambda_{\max}^B D_z^2 \sqrt{T}/2 + \sqrt{2} G_f D_x \sqrt{T}/\sqrt{\alpha}$$

$$R^C(T) \leq \lambda_{\max}^B D_z^2 + 2\sqrt{2} D_x G_f/\sqrt{\alpha} + 2F\sqrt{T}$$

- No assumptions on **A**, **B**, ***c***, and the subgradient of *g* (suits indicator fn.)
- If $\|\mathbf{y}_t\| \leq D, \ \forall t$, then

$$\sum_{t=1}^{T} \|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c}\|^2 \leq 4D^2$$

- Proof: Based on Lemma 1, three point property of Bergman divergence, and Fenchel-young inequality ….

Proof sketch

$$-\mathbf{A}^\top(\mathbf{y}_{t+1} + \rho\mathbf{B}(\mathbf{z}_t - \mathbf{z}_{t+1})) - \eta(\nabla\phi(\mathbf{x}_{t+1}) - \nabla\phi(\mathbf{x}_t)) \in \partial f_t(\mathbf{x}_{t+1})$$

- From Lemma 1

$$\begin{aligned}\left[f_t(\mathbf{x}_{t+1}) + g(\mathbf{z}_{t+1})\right] - \left[f_t(\mathbf{x}^*) + g(\mathbf{z}^*)\right] \leq{}& \frac{1}{2\rho}\left(\|\mathbf{y}_t\|^2 - \|\mathbf{y}_{t+1}\|^2\right) \\ & -\frac{\rho}{2}\|\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_t - \mathbf{c}\|^2 + \frac{\rho}{2}\left(\|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_t)\|^2 - \|\mathbf{B}(\mathbf{z}^* - \mathbf{z}_{t+1})\|\right) \\ & -\eta\langle\nabla\phi(\mathbf{x}_{t+1}) - \nabla\phi(\mathbf{x}_t), \mathbf{x}_{t+1} - \mathbf{x}^*\rangle\end{aligned}$$

- Three point property

$$-\langle\nabla\phi(\mathbf{x}_{t+1}) - \nabla\phi(\mathbf{x}_t), \mathbf{x}_{t+1} - \mathbf{x}^*\rangle = B_\phi(\mathbf{x}^*, \mathbf{x}_t) - B_\phi(\mathbf{x}^*, \mathbf{x}_{t+1}) - B_\phi(\mathbf{x}_{t+1}, \mathbf{x}_t)$$

- Fenchel-Young's inequality

$$f_t(\mathbf{x}_t) - f_t(\mathbf{x}_{t+1}) \leq \langle f_t'(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t+1}\rangle \leq \tfrac{1}{2\alpha\eta}\|f_t'(\mathbf{x}_t)\|_q^2 + \tfrac{\alpha\eta}{2}\|\mathbf{x}_t - \mathbf{x}_{t+1}\|_p^2$$

- Finally, bounding

$$\left[f_t(\mathbf{x}_t) + g(\mathbf{z}_{t+1})\right] - \left[f_t(\mathbf{x}^*) + g(\mathbf{z}^*)\right]$$

and summing up both sides the result follows.

Strongly convex f_t, g ($\eta > 0$)

- f_t : β_1-strongly convex

$$f_t(\mathbf{x}^*) \geq f_t(\mathbf{x}) + \langle f_t^{'}(\mathbf{x}), \mathbf{x}^* - \mathbf{x} \rangle + \beta_1 B_\Phi(\mathbf{x}^*, \mathbf{x})$$

- g: β_2-strongly convex

$$g_t(\mathbf{z}^*) \geq g(\mathbf{z}) + \langle g^{'}(\mathbf{z}), \mathbf{z}^* - \mathbf{z} \rangle + \frac{\beta_2}{2} \|\mathbf{x}^* - \mathbf{x}\|^2$$

Theorem 3: If (a1) – (a4) hold, then for $\eta_t = \beta_1 t$ and $\rho_t = \beta_2 t / \lambda_{\max}^B$

$$R_1(T) \leq \frac{G_f^2}{2\alpha\beta_1} \log(T+1) + \frac{\beta_2 D_z^2}{2} + \beta_1 D_x^2$$

$$R^C(T) \leq \frac{2F\lambda_{\max}^B}{\beta_2} \log(T+1) + \lambda_{\max}^B D_z^2 + \frac{2\beta_1 \lambda_{\max}^B D_x^2}{\beta_2}$$

OADM with $\eta = 0$

- Algorithm

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ f_t(\mathbf{x}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \rangle + \frac{\rho}{2} \| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \|^2 \},$$

$$\mathbf{z}_{t+1} = \arg\min_{\mathbf{z}} \{ g(\mathbf{z}) + \langle \mathbf{y}_t, \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2} \| \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z} - \mathbf{c} \|^2 \},$$

$$\mathbf{y}_{t+1} = \mathbf{y}_t + \rho(\mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{z}_{t+1} - \mathbf{c})$$

- $(\hat{\mathbf{x}}_t, \mathbf{z}_t)$ s.t. $\mathbf{A}\hat{\mathbf{x}}_t + \mathbf{B}\mathbf{z}_t = \mathbf{c}$ [e.g., in consensus opt. $\hat{\mathbf{x}}_t = \mathbf{z}_t$]
- New assumptions: (a5) **A** square and invertible, (a6) $\|f'(\mathbf{x})\|_2 \leq G_f$
- Regret of $\{\hat{\mathbf{x}}_t, \mathbf{z}_t\}_{t=1}^T$

$$R_2(T) := \sum_{t=1}^T (f_t(\hat{\mathbf{x}}_t) + g(\mathbf{z}_t)) - \min_{\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{z}=\mathbf{c}} \sum_{t=1}^T (f_t(\mathbf{x}) + g(\mathbf{z}))$$

Regret bounds

Theorem 4: If (a1) – (a6) hold, and $\rho = \dfrac{G_f\sqrt{T}}{D_z\sqrt{\lambda^A_{\min}\lambda^B_{\max}}}$, then

$$R_2(T) \le G_f D_z \sqrt{\frac{\lambda^B_{\max}}{\lambda^A_{\min}}}\sqrt{T}$$

$$R^C(T) \le \lambda^B_{\max} D_z^2 + \frac{2FD_z\sqrt{\lambda^B_{\max}\lambda^A_{\min}T}}{G_f}$$

Theorem 5: Under (a1)-(a6), if *g* is β_2-strongly convex, and $\rho_t = \beta_2 t/\lambda^B_{\max}$,

$$R_2(T) \le \frac{G_f^2}{2\beta_2}\frac{\lambda^B_{\max}}{\lambda^A_{\min}}\log(T+1) + \beta_2 D_z^2$$

$$R^C(T) \le \lambda^B_{\max} D_z^2 + \frac{2F\lambda^B_{\max}}{\beta_2}\log(T+1)$$

Inexact OADM

- Expensive to solve for $\mathbf{x}_t$, exactly, e.g., logistic regression loss
- Theorems 2, 3 still hold for:

Case 1) Linearizing f_t

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ \langle f_t'(\mathbf{x}_t), \mathbf{x} - \mathbf{x}_t \rangle + \frac{\rho}{2} \| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_t - \mathbf{c} \|^2 + \eta B_\phi(\mathbf{x}, \mathbf{x}_t) \}$$

Case 2) Linearizing both f_t and quadratic penalty

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ \langle F_t(\mathbf{x}_t), \mathbf{x} \rangle + \eta B_\psi(\mathbf{x}, \mathbf{x}_t) \}$$

Case 3) Composite objective $f_t = f_t^S + f_t^N$ (COMID)

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \{ f_t^N(\mathbf{x}) + \langle F_t^S(\mathbf{x}_t), \mathbf{x} - \mathbf{x}_t \rangle + \eta B_\psi(\mathbf{x}, \mathbf{x}_t) \}$$

Stochastic OADM

- Stochastic formulation: $\min_{\mathbf{x},\mathbf{y}} \ \mathbb{E}_\xi[f(\mathbf{x},\xi)] + g(\mathbf{z}) \quad \text{s.to} \ \ \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$
- $f(\mathbf{x}) := \mathbb{E}_\xi[f(\mathbf{x},\xi)]$ → $f'(\mathbf{x}_t,\xi_t)$: unbiased estimate of $f'(\mathbf{x}_t)$
- $\{\xi_t\}_{t=1}^T$ → Inexact OADM → $\{\mathbf{x}_t,\mathbf{y}_t,\mathbf{z}_t\}_{t=1}^T$

Corollary 1: Under (a1) – (a3), if $\eta = \frac{G_f\sqrt{T}}{D_x\sqrt{2\alpha}}$ and $\rho = \sqrt{T}$, then

(a) Expected regret

$$\mathbb{E}[\underbrace{f(\bar{\mathbf{x}}_T) + g(\bar{\mathbf{z}}_T)}_{\Theta_1}] \leq \underbrace{f(\mathbf{x}^*) + g(\mathbf{z}^*) + \frac{\lambda_{\max}^B D_z^2}{2\sqrt{T}} + \frac{\sqrt{2}G_f D_x}{\sqrt{\alpha T}}}_{\mu_1}$$

$$\mathbb{E}[\|\underbrace{\mathbf{A}\bar{\mathbf{x}}_T + \mathbf{B}\bar{\mathbf{z}}_T - \mathbf{c}}_{\Theta_C}\|^2] \leq \underbrace{\frac{\lambda_{\max}^B D_z^2}{T} + \frac{2\sqrt{2}D_x G_f}{\sqrt{\alpha T}} + \frac{2F}{\sqrt{T}}}_{\mu_C}$$

(b) High-probability regret

$$\mathrm{P}[\Theta_1 - \mu_1 \geq \epsilon], \ \mathrm{P}[\Theta_C - \mu_C \geq \epsilon] \leq \exp\left(-\frac{T\alpha\epsilon^2}{16D_x^2G_f^2}\right)$$