Robust Convergence Analysis of Three-Operator Splitting

Han Wang, Mahyar Fazlyab, Shaoru Chen, Victor M. Preciado

Department of Electrical and Systems Engineering
University of Pennsylvania

September 25, 2019

Outline

Motivation

Optimization problems of the form

$$\min_{x \in \mathbb{R}^d} F(x) = f(x) + g(x) + h(x)$$

arise in many applications such as

- Control: optimal control problem
- Signal processing: image inpainting
- Machine learning: group LASSO, support vector machine

Applications

Box-constrained optimal control problem

$$
\begin{aligned}
\underset{x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^m}{\text{minimize}} \quad & \frac{1}{2}\Big(\sum_{t=0}^{N} x_t^T Q_t x_t + \sum_{t=0}^{N-1} u_t^T R_t u_t\Big) \\
\text{subject to} \quad & x_{t+1} = A_t x_t + B_t u_t, \quad t = 0, \cdots, N-1 \\
& \|u_t\|_\infty \leq 1, \quad t = 0, \cdots, N-1 \\
& x_0 = x_{\text{init}}
\end{aligned}
$$

Concatenate states x_t and inputs u_t as

$$
x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad u = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-1} \end{bmatrix}, \quad (x, u) = \begin{bmatrix} x \\ u \end{bmatrix}.
$$

Rewrite the optimal control problem as a three-operator splitting problem

$$\underset{(x,u)}{\text{minimize}} \quad I_{\mathcal{C}}((x,u)) + I_{\mathcal{D}}((x,u)) + \frac{1}{2}(x,u)^T E(x,u)$$

with indicator functions $I_{\mathcal{C}}, I_{\mathcal{D}}$ encoding the constraints

$$\mathcal{D} = \{(x,u) \mid x_0 = x_{\text{init}}, x_{t+1} = A_t x_t + B_t u_t, t = 0, \cdots, N-1\},$$
$$\mathcal{C} = \{(x,u) \mid \|u_t\|_\infty \leq 1, t = 0, \cdots, N-1\},$$
$$E = \text{diag}(Q_0, \cdots, Q_N, R_0, \cdots, R_{N-1}).$$

Image inpainting problem [Davis and Yin, 2015]

$$\underset{\mathbf{x}}{\text{minimize}} \quad \omega\|\mathbf{x}_{(1)}\|_* + \omega\|\mathbf{x}_{(2)}\|_* + \frac{1}{2}\|P_\Omega \mathbf{x} - P_\Omega \mathbf{y}\|^2$$

where $\mathbf{y}, \mathbf{x}$ are 3-way tensors representing original and recovered images. P_Ω : observation filter. $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}$: row and column concatenations of layers of $\mathbf{x}$.

Example: occluded image $\rightarrow$ recovered image.

Three-operator splitting algorithm (TOS)

- [Davis and Yin, 2015] proposed the TOS algorithm

$$\min_{x \in \mathbb{R}^d} F(x) = f(x) + g(x) + h(x)$$

Algorithm 1 Three-Operator Splitting (TOS)

Input: $z_0 \in \mathbb{R}^d, \alpha, \lambda > 0$.

for $k = 0, 1, 2, \cdots$

$$\begin{aligned} x_B^k &= \mathrm{prox}_{\alpha g}(z^k); \\ y^k &= 2x_B^k - z^k - \alpha \nabla h(x_B^k); \\ x_A^k &= \mathrm{prox}_{\alpha f}(y^k); \\ z^{k+1} &= z^k + \lambda(x_A^k - x_B^k); \end{aligned}$$

endfor

---

where

$$\mathrm{prox}_f(x) = \mathrm{argmin}_y \; f(y) + \frac{1}{2}\|x - y\|^2$$

Three-operator splitting extensions

- Adaptive TOS [Pedregosa and Gidel, 2018]
- Momentum accelerated TOS [Davis and Yin, 2015]
- Stochastic TOS [Yurtsever et al., 2016]
- Inexact TOS [Zong et al., 2018]

Our contribution

Provide a **unified framework** for convergence analysis and parameter selection of the TOS algorithm using robust control theory and semi-definite programming.

Our approach

Penn
University of Pennsylvania

1. View the algorithm as a linear system with uncertainty ($\text{prox}_{\alpha g}, \text{prox}_{\alpha f}, \nabla h$) and use incremental quadratic constraints to model the uncertainties.

2. Design a Lyapunov function to certify convergence under various assumptions about f, g, and h.

3. Use S-procedure to find sufficient conditions for Lyapunov stability in terms of linear matrix inequalities (LMIs).

Related work

- Lyapunov/control theoretic methods in optimization: [Wang and Elia, 2010, Wang and Elia, 2011, Su et al., 2014, Wilson et al., 2016, Wibisono et al., 2016] and many others
- Worst-case convergence analysis using semi-definite programming: [Drori and Teboulle, 2014, Lessard et al., 2016, Fazlyab et al., 2018, Taylor et al., 2017, Ryu et al., 2018, Seidman et al., 2019] and many others

Main results

Notation: $f \in \mathcal{F}(m_f, L_f)$ if f is Lipschitz differentiable with Lipschitz constant $0 \leq L_f < \infty$ and f is strongly convex with parameter $0 \leq m_f \leq L_f$ ($L_f = \infty$ when non-differentiable)

Assumption: $f \in \mathcal{F}(m_f, L_f)$, $g \in \mathcal{F}(m_g, L_g)$, $h \in \mathcal{F}(m_h, L_h)$

Case 1: $L_h < \infty, L_f = L_g = \infty$, $m_f = m_g = m_h = 0$

$$\min_{i=0,\cdots,k-1} \|\partial f(x_A^i) + \partial g(x_B^i) + \nabla h(x_B^i)\|^2 \leq \frac{\|z^0 - z^\star\|^2}{\theta k}$$

Case 2: $L_f, L_h < \infty, L_g = \infty$, $m_f = m_g = m_h = 0$

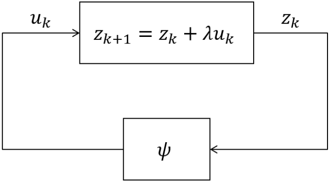
$$\min_{i=0,\cdots,k-1} [F(x_B^i) - F(x_B^\star)] \leq \frac{1}{\theta k}\|z^0 - z^\star\|^2.$$

Case 3: $(m_f + m_g + m_h)(1/L_f + 1/L_g)1/L_h > 0$

$$\|z^k - z^\star\|^2 \leq \rho^{2k}\|z^0 - z^\star\|^2.$$

Robust control perspective

We can view the TOS algorithm as a linear system + nonlinear feedback



where

$$\psi(z^k) = \quad \mathrm{prox}_{\alpha f}(2\mathrm{prox}_{\alpha g}(z^k) - z_k - \alpha \nabla h(\mathrm{prox}_{\alpha g}(z^k))) - \mathrm{prox}_{\alpha g}(z^k).$$

Problem: show Lyapunov stability and certify worst-case convergence rates for all $f \in \mathcal{F}(m_f, L_f), g \in \mathcal{F}(m_g, L_g)$, and $h \in \mathcal{F}(m_h, L_h)$

If f is differentiable, then $f \in \mathcal{F}(m, L)$ if and only if [Nesterov, 1998]

$$\begin{bmatrix} x - y \\ \nabla f(x) - \nabla f(y) \end{bmatrix}^T Q(m, L) \begin{bmatrix} x - y \\ \nabla f(x) - \nabla f(y) \end{bmatrix} \geq 0 \qquad (1)$$

for $\forall x, y \in \mathrm{dom} f$ where

$$Q(m, L) = \begin{bmatrix} -\frac{mL}{m+L} & 1/2 \\ 1/2 & -\frac{1}{m+L} \end{bmatrix} \otimes I_d.$$

Eq. (1) is called an incremental quadratic constraint defined by $Q(m, L)$.
In this way, we can characterize all $f \in \mathcal{F}(m_f, L_f)$ through (1).

Case 1

Case 1: $f \in \mathcal{F}(0,\infty)$, $g \in \mathcal{F}(0,\infty)$, $h \in \mathcal{F}(0,L_h)$, $L_h < \infty$

Define the following Lyapunov function

$$V_k = \|z^k - z^\star\|^2 + \theta \sum_{i=0}^{k-1} \|\partial f(x_A^i) + \partial g(x_B^i) + \nabla h(x_B^i)\|^2$$

with a tuning parameter $\theta > 0$. If $V_{k+1} \leq V_k \Rightarrow$

$$\min_{i=0,\cdots,k-1} \|\partial f(x_A^i) + \partial g(x_B^i) + \nabla h(x_B^i)\|^2 \leq \frac{\|z^0 - z^\star\|^2}{\theta k}.$$

Main result: case 1

Theorem

Let $m_f = m_g = m_h = 0$ and $L_h < L_f = L_g = \infty$. Suppose there exist $\lambda, \alpha, \theta > 0, \sigma_1, \sigma_2, \sigma_3 \geq 0$ such that the following matrix inequality

$$W_0 + \sigma_1 Q_1 + \sigma_2 Q_2 + \sigma_3 Q_3 \preceq 0$$

holds, then for all f, g which are convex,proper and closed, h which is L_h-Lipschitz differentiable, the TOS algorithm satisfies

$$\min_{i=0,\cdots,k-1} \|\partial f(x_A^i) + \partial g(x_B^i) + \nabla h(x_B^i)\|^2 \leq \frac{\|z^0 - z^\star\|^2}{\theta k}$$

Main result: case 1

$$W_0 = \begin{bmatrix} \lambda^2+\theta/\alpha^2 & 0 & -\lambda^2-\theta/\alpha^2 & -\lambda \\ 0 & 0 & 0 & 0 \\ -\lambda^2-\theta/\alpha^2 & 0 & \lambda^2+\theta/\alpha^2 & \lambda \\ -\lambda & 0 & \lambda & 0 \end{bmatrix} \otimes I_d,$$

$$Q_1 = \begin{bmatrix} \alpha I_d & -I_d \\ 0 & 0 \\ 0 & 0 \\ 0 & I_d \end{bmatrix} Q(m_g, L_g) \begin{bmatrix} \alpha I_d & 0 & 0 & 0 \\ -I_d & 0 & 0 & I_d \end{bmatrix},$$

$$Q_2 = \begin{bmatrix} \alpha I_d & 2I_d \\ 0 & -I_d \\ 0 & 0 \\ 0 & -I_d \end{bmatrix} Q(m_h, L_h) \begin{bmatrix} \alpha I_d & 0 & 0 & 0 \\ 2I_d & -I_d & 0 & -I_d \end{bmatrix},$$

$$Q_3 = \begin{bmatrix} 0 & 0 \\ 0 & I_d \\ \alpha I_d & -I_d \\ 0 & 0 \end{bmatrix} Q(m_f, L_f) \begin{bmatrix} 0 & 0 & \alpha I_d & 0 \\ 0 & I_d & -I_d & 0 \end{bmatrix}.$$

Case 1

Define

$$v_k = \left[(x_B^k - x_B^\star)^\top \ (y^k - y^\star)^\top \ (x_A^k - x_A^\star)^\top \ (z^k - z^\star)^\top \right]^\top.$$

Then $V_{k+1} - V_k = v_k^\top W_0 v_k$ where

$$W_0 = \begin{bmatrix} \lambda^2 + \theta/\alpha^2 & 0 & -\lambda^2 - \theta/\alpha^2 & -\lambda \\ 0 & 0 & 0 & 0 \\ -\lambda^2 - \theta/\alpha^2 & 0 & \lambda^2 + \theta/\alpha^2 & \lambda \\ -\lambda & 0 & \lambda & 0 \end{bmatrix} \otimes I_d.$$

Subproblem: Find $\alpha, \lambda, \theta > 0$ such that $W_0 \preceq 0$ for all $f \in \mathcal{F}(m_f, L_f)$, $g \in \mathcal{F}(m_g, L_g)$, $h \in \mathcal{F}(m_h, L_h)$.

Case 1

By incremental quadratic constraints, we can define Q_1, Q_2, Q_3 such that

$$f \in \mathcal{F}(m_f, L_f) \Rightarrow v_k^\top Q_1 v_k \geq 0$$
$$h \in \mathcal{F}(m_h, L_h) \Rightarrow v_k^\top Q_2 v_k \geq 0$$
$$g \in \mathcal{F}(m_g, L_g) \Rightarrow v_k^\top Q_3 v_k \geq 0$$

Using S-procedure, we obtain the following sufficient condition: If there exist $\alpha, \lambda, \theta > 0$, $\sigma_1, \sigma_2, \sigma_3 \geq 0$ such that

$$W_0 + \sigma_1 Q_1 + \sigma_2 Q_2 + \sigma_3 Q_3 \preceq 0,$$

then the Lypaunov function V_k is non-increasing and the TOS algorithm achieves sublinear convergence (worst-case convergence rate).

Case 1

In this case, when $m_f = m_g = m_h = 0, L_h < L_f = L_g = \infty$, the matrix inequality

$$W_0 + \sigma_1 Q_1 + \sigma_2 Q_2 + \sigma_3 Q_3 \preceq 0$$

can be solved symbolically by

$$\alpha = (2-\lambda)/L_h, \quad \theta = (2-\lambda)^3\lambda/(2L_h^2) = \alpha^2\lambda(2-\lambda)/2,$$
$$\sigma_1 = \sigma_2 = \sigma_3 = \frac{2\lambda}{\alpha}$$

for $\lambda \in (0,2)$.

Parameter Selection: maximize θ to achieve the best convergence rate. $\theta = \frac{27}{32L_h^2}$ when $\lambda^\star = \frac{1}{2}$.

Case 3

Under assumption $(m_f + m_g + m_h)(1/L_f + 1/L_g)1/L_h > 0$, the TOS algorithm achieves linear convergence rate [Davis and Yin, 2015], i.e.,

$$\|z^k - z^\star\|^2 \leq \rho^{2k}\|z^0 - z^\star\|^2.$$

We define the Lyapunov function as

$$V_k = \|z^k - z^\star\|^2.$$

Then, $V_{k+1} - \rho^2 V_k = v_k^T W_1 v_k$ with

$$W_1 = \begin{bmatrix} \lambda^2 & 0 & -\lambda^2 & -\lambda \\ 0 & 0 & 0 & 0 \\ -\lambda^2 & 0 & \lambda^2 & \lambda \\ -\lambda & 0 & \lambda & 1-\rho^2 \end{bmatrix} \otimes I_d.$$

Case 3

Similarly, if there exists $\lambda, \alpha > 0, \sigma_1, \sigma_2, \sigma_3 \geq 0, 0 < \rho^2 < 1$ such that

$$W(\alpha, \lambda, \rho^2, \sigma_1, \sigma_2, \sigma_3) = W_1 + \sigma_1 Q_1 + \sigma_2 Q_2 + \sigma_3 Q_3 \preceq 0 \qquad (3)$$

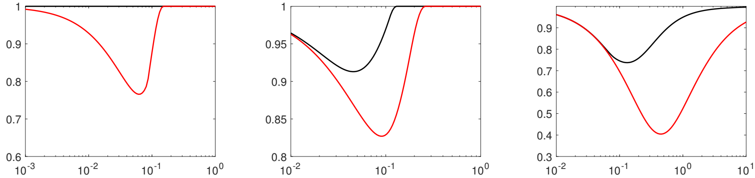
is feasible for some $\rho^2 < 1 \Rightarrow$ linear convergence.

- For fixed λ, α, (3) becomes an LMI in $\rho^2, \sigma_1, \sigma_2, \sigma_3$.
- We can convexify the dependence of W on λ by Schur Complements (more details in the paper!).

Case 3

Let

$$\rho^\star = \min_{\alpha > 0} \rho^\star(\alpha).$$



(a) $L_f = 50, m_h = 2, L_h = 30$

(b) $m_f = 20, L_f = 20, L_h = 70$

(c) $m_f = 1, L_g = 5, L_h = 1/9$

Figure: $\rho^\star(\alpha)^2 - \alpha$ plots. Red: our results; Black: results in [Ryu et al., 2018].

Case 2: $f \in \mathcal{F}(0, L_f)$, $g \in \mathcal{F}(0, \infty)$, $h \in \mathcal{F}(0, L_h)$, $L_h, L_f < \infty$

Define the Lyapunov function

$$V_k = \|z^k - z^\star\|^2 + \theta \sum_{i=0}^{k-1} [F(x_B^i) - F(x_B^\star)].$$

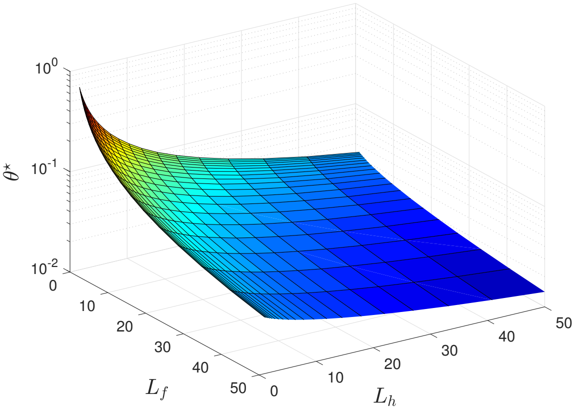
Similarly, we can define a matrix inequality such that any feasible solution $(\alpha, \lambda, \theta, \sigma_1, \sigma_2, \sigma_3)$ to

$$W_2 + \sigma_1 Q_1 + \sigma_2 Q_2 + \sigma_3 Q_3 \preceq 0$$

leads to $V_{k+1} \leq V_k$ and certifies the sublinear convergence rate of the TOS algorithm as

$$\min_{i=0,\cdots,k-1} [F(x_B^i) - F(x_B^\star)] \leq \frac{1}{\theta k} \|z^0 - z^\star\|^2.$$

Case 2



Largest $\theta^\star$ found by SDPs

Optimal control problem revisited

Recall the box-constrained optimal control problem
Box-constrained optimal control problem

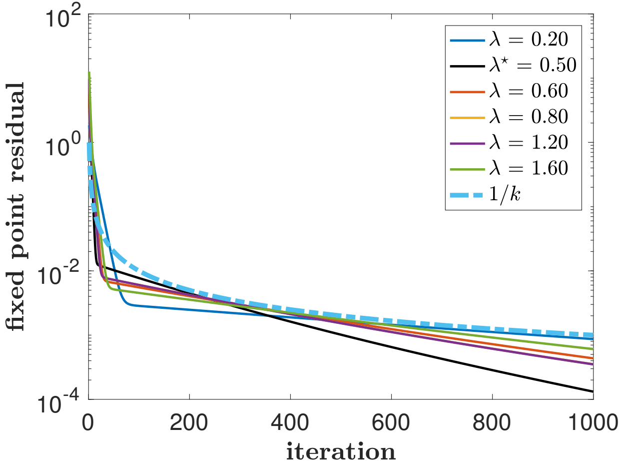
$$
\begin{array}{ll}
\underset{x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^m}{\text{minimize}} & \frac{1}{2}\left(\sum_{t=0}^{N} x_t^T Q_t x_t + \sum_{t=0}^{N-1} u_t^T R_t u_t\right) \\
\text{subject to} & x_{t+1} = A_t x_t + B_t u_t, \quad t = 0, \cdots, N-1 \\
& \|u_t\|_\infty \leq 1, \quad t = 0, \cdots, N-1 \\
& x_0 = x_{\text{init}}
\end{array}
$$

Setup: $x_t \in \mathbb{R}^{20}$, $u_t \in \mathbb{R}^5$, $N = 20$. Use the TOS algorithm to solve

$$
\underset{(x,u)}{\text{minimize}} \quad I_{\mathcal{C}}((x,u)) + I_{\mathcal{D}}((x,u)) + \frac{1}{2}(x,u)^T E (x,u)
$$

From the analysis of Case 1 ($L_h < \infty$), $\lambda = \frac{1}{2}$ gives the worst-case optimal convergence rate.

Optimal control problem revisited



Take away

1. We proposed a unified framework to derive worst-case convergence rates for the TOS algorithm under various assumptions.

2. The main tools are robust control theory and semi-definite programming.

3. The framework provides guidelines to algorithm parameter selection.

Robust Convergence Analysis of Three-Operator Splitting

Han Wang, Mahyar Fazlyab, Shaoru Chen, Victor M. Preciado

Department of Electrical and Systems Engineering
University of Pennsylvania

September 25, 2019

Davis, D. and Yin, W. (2015).
A three-operator splitting scheme and its optimization applications.

Drori, Y. and Teboulle, M. (2014).
Performance of first-order methods for smooth convex minimization: a novel approach.
Mathematical Programming, 145(1-2):451–482.

Fazlyab, M., Ribeiro, A., Morari, M., and Preciado, V. M. (2018).
Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems.
SIAM Journal on Optimization, 28(3):2654–2689.

Lessard, L., Recht, B., and Packard, A. (2016).
Analysis and design of optimization algorithms via integral quadratic constraints.

SIAM Journal on Optimization, 26(1):57–95.

Nesterov, Y. (1998).
Introductory lectures on convex programming.

Pedregosa, F. and Gidel, G. (2018).
Adaptive three operator splitting.
arXiv preprint arXiv:1804.02339.

Ryu, E. K., Taylor, A. B., Bergeling, C., and Giselsson, P. (2018).
Operator splitting performance estimation: Tight contraction factors and optimal parameter selection.
arXiv preprint arXiv:1812.00146.

Seidman, J. H., Fazlyab, M., Preciado, V. M., and Pappas, G. J. (2019).
A control-theoretic approach to analysis and parameter selection of douglas–rachford splitting.

IEEE Control Systems Letters, 4(1):199–204.

Su, W., Boyd, S., and Candes, E. (2014).
A differential equation for modeling nesterovâs accelerated gradient method: Theory and insights.
In *Advances in Neural Information Processing Systems*, pages 2510–2518.

Taylor, A. B., Hendrickx, J. M., and Glineur, F. (2017).
Exact worst-case performance of first-order methods for composite convex optimization.
SIAM Journal on Optimization, 27(3):1283–1313.

Wang, J. and Elia, N. (2010).
Control approach to distributed optimization.
In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 557–561. IEEE.

Wang, J. and Elia, N. (2011).

A control perspective for centralized and distributed convex optimization.
In *2011 50th IEEE conference on decision and control and European control conference*, pages 3800–3805. IEEE.

Wibisono, A., Wilson, A. C., and Jordan, M. I. (2016).
A variational perspective on accelerated methods in optimization.
proceedings of the National Academy of Sciences, 113(47):E7351–E7358.

Wilson, A. C., Recht, B., and Jordan, M. I. (2016).
A lyapunov analysis of momentum methods in optimization.
arXiv preprint arXiv:1611.02635.

Yurtsever, A., Vu, B. C., and Cevher, V. (2016).
Stochastic three-composite convex minimization.

In *Advances in Neural Information Processing Systems*, pages 4329–4337.

Zong, C., Tang, Y., and Cho, Y. (2018).
Convergence analysis of an inexact three-operator splitting algorithm.
Symmetry, 10(11):563.