Contraction Theory for Networked Optimization and Control

Francesco Bullo

Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara
https://fbullo.github.io/ctds

Workshop on Emerging Challenges of Network-Enabled Control and Optimization
CDC Workshop, Singapore, Dec 12, 2023

Many thanks to: General Chair Jie Chen and General Co-Chair Huanshui Zhang
Shandong University of Science and Technology & Nanyang Technological University

Acknowledgments

Veronica Centorrino
Scuola Sup Meridionale

Lily Cothren
UC Boulder

Emiliano Dall'Anese
UC Boulder

Alex Davydov
UC Santa Barbara

Anand Gokhale
UC Santa Barbara

Giovanni Russo
Univ Salerno

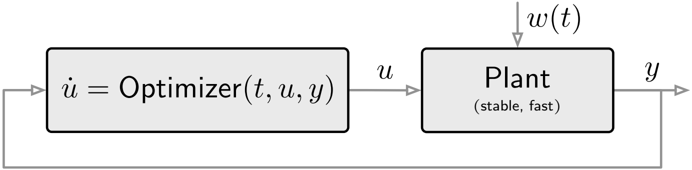
A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization. *IEEE Transactions on Automatic Control*, June 2023. . Submitted

A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, Sept. 2023. . To appear

L. Cothren, F. Bullo, and E. Dall'Anese. Singular perturbation via contraction theory. *IEEE Transactions on Automatic Control*, Oct. 2023. . Submitted

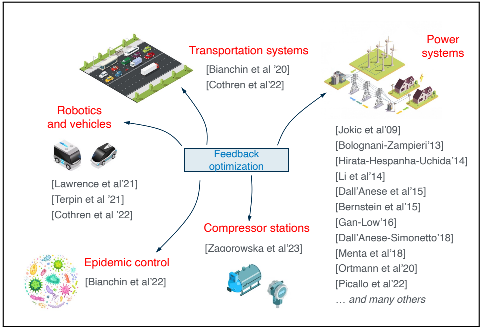
Solving optimization problems via dynamical systems

- studies in linear and nonlinear programming (Arrow, Hurwicz, and Uzawa 1958)
- neural networks (Hopfield and Tank 1985)
- analog circuits (Kennedy and Chua 1988)
- optimization on manifolds (Brockett 1991)
- ...



Continuous-time optimizations solvers

1. **online feedback optimization**
2. distributed optimization
3. parametric convex optimization
4. model predictive control
5. control barrier functions

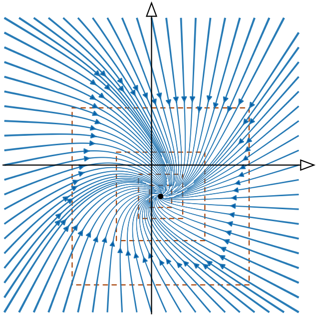


Online feedback optimization

Slide courtesy of Emiliano Dall'Anese, University of Colorado Boulder

contractivity = robust computationally-friendly stability

fixed point theory + Lyapunov stability theory + geometry of metric spaces

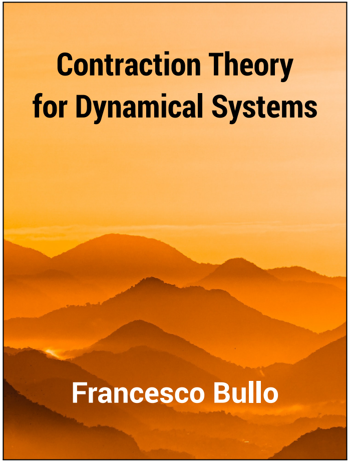


search for contraction properties

design engineering systems to be contracting

verify correct/safe behavior via known Lipschitz constants

Recent education and research on Contraction Theory



"Continuous improvement is better than delayed perfection" **Mark Twain**

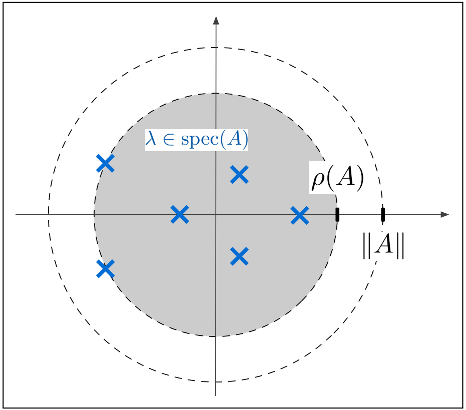
- 2023 ACC Workshop "Contraction Theory for Systems, Control, and Learning" **http://motion.me.ucsb.edu/contraction-workshop-2023**
- 2021 IEEE CDC Tutorial session "Contraction Theory for Machine Learning" **https://sites.google.com/view/contractiontheory** (PDFs and youtube videos)
- 2022 IEEE CDC plenary presentation "Contraction Theory in Systems and Control" **https://fbullo.github.io/talks/2022-12-FBullo-ContractionSystemsControl-CDC.pdf**
- Textbook: Contraction Theory for Dynamical Systems, Francesco Bullo, rev 1.1, Mar 2023. (Book and slides freely available) **https://fbullo.github.io/ctds**
- 2023 Comprehensive tutorial slides: **https://fbullo.github.io/ctds**
- 2023 Sep: Youtube lectures: "Minicourse on Contraction Theory" **https://youtu.be/FQV5PrRHks8** 12h in 6 lectures

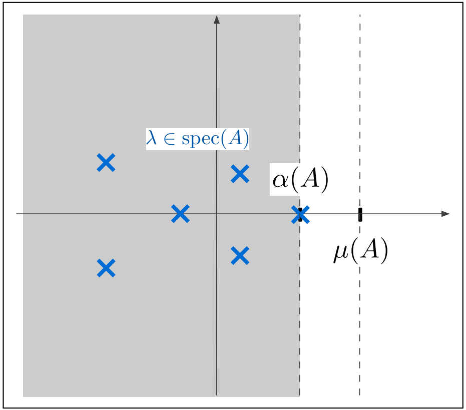
- Three CDC2023 invited sessions on **Contraction Theory for Analysis, Synchronization and Regulation**, tomorrow Wednesday!

Outline

Induced matrix norms

Vector norm	Induced matrix norm	Induced matrix log norm
$\|x\|_1 = \sum_{i=1}^{n} \|x_i\|$	$\|A\|_1 = \max_{j\in\{1,\dots,n\}} \sum_{i=1}^{n} \|a_{ij}\|$ $=$ max column "absolute sum" of A	$\mu_1(A) = \max_{j\in\{1,\dots,n\}} \left(a_{jj} + \sum_{i=1,i\neq j}^{n} \|a_{ij}\|\right)$ absolute value only off-diagonal
$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$	$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$	$\mu_2(A) = \lambda_{\max}\left(\frac{A + A^\top}{2}\right)$
$\|x\|_\infty = \max_{i\in\{1,\dots,n\}} \|x_i\|$	$\|A\|_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^{n} \|a_{ij}\|$ $=$ max row "absolute sum" of A	$\mu_\infty(A) = \max_{i\in\{1,\dots,n\}} \left(a_{ii} + \sum_{j=1,j\neq i}^{n} \|a_{ij}\|\right)$ absolute value only off-diagonal


$\lambda \in \mathrm{spec}(A)$
$\rho(A)$
$\|A\|$


$\lambda \in \mathrm{spec}(A)$
$\alpha(A)$
$\mu(A)$

Continuous-time dynamics and one-sided Lipschitz constants

$$\dot{x} = \mathsf{F}(x) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\| \text{ and induced log norm } \mu(\cdot)$$

One-sided Lipschitz constant

$$\begin{aligned}\mathrm{osLip}(\mathsf{F}) &= \inf\{b \in \mathbb{R} \text{ such that } [\![\mathsf{F}(x) - \mathsf{F}(y), x - y]\!] \leq b\|x - y\|^2 \quad \text{for all } x, y\} \\ &= \sup_x \mu(D\mathsf{F}(x))\end{aligned}$$

For **scalar map** f, $\quad \mathrm{osLip}(f) = \sup_x f'(x)$

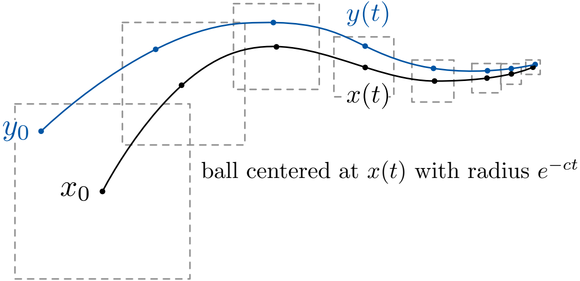
For **affine map** $\mathsf{F}_A(x) = Ax + a$

$$\begin{aligned}\mathrm{osLip}_{2,P}(\mathsf{F}_A) = \mu_{2,P}(A) \leq \ell \quad &\iff \quad A^\top P + AP \preceq 2\ell P \\ \mathrm{osLip}_{\infty,\eta}(\mathsf{F}_A) = \mu_{\infty,\eta}(A) \leq \ell \quad &\iff \quad a_{ii} + \sum_{j \neq i} |a_{ij}|\eta_i/\eta_j \leq \ell\end{aligned}$$

Banach contraction theorem for continuous-time dynamics:

If $-c := \text{osLip}(\mathsf{F}) < 0$, then

1. F is **infinitesimally contracting** = distance between trajectories decreases exp fast (e^{-ct})
2. F has a unique, glob exp stable equilibrium x^*
3. global Lyapunov functions $V_1(x) = \|x - x^*\|^2$ and $V_2(x) = \|\mathsf{F}(x)\|^2$



Euler discretization theorem for contracting dynamics

Given arbitrary norm $\|\cdot\|$ and differentiable $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, equivalent statements

1. $\dot{x} = \mathsf{F}(x)$ is infinitesimally contracting
2. there exists $\alpha > 0$ such that $x_{k+1} = x_k + \alpha \mathsf{F}(x_k)$ is contracting

Interconnected subsystems: $x_i \in \mathbb{R}^{N_i}$ and $x_{-i} \in \mathbb{R}^{N-N_i}$:

$$\dot{x}_i = \mathsf{F}_i(x_i, x_{-i}), \qquad \text{for } i \in \{1, \dots, n\}$$

Network contraction theorem. Assume

- **contractivity wrt x_i**: $\quad \mathrm{osLip}_{x_i}(\mathsf{F}_i) \leq -c_i < 0, \quad$ uniformly in x_{-i}
- **Lipschitz wrt x_j, $j \neq i$**: $\quad \mathrm{Lip}_{x_j}(\mathsf{F}_i) \leq \ell_{ij}, \quad$ uniformly in x_{-j}
- the Lipschitz constants matrix $\Gamma = \begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$ is **Hurwitz**

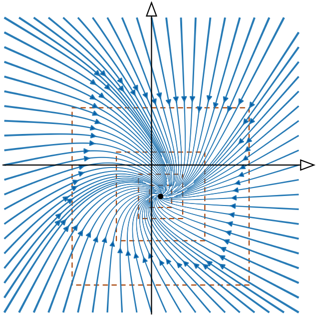
$\Longrightarrow$ **interconnected system** is contracting wrt rate $|\alpha(\Gamma)|$

contractivity = robust computationally-friendly stability

fixed point theory + Lyapunov stability theory + geometry of metric spaces

highly-ordered transient and asymptotic behavior, no anonymous constants/functions:

1. unique globally exponential stable equilibrium
   & two natural Lyapunov functions
2. robustness properties
   - bounded input, bounded output (iss)
   - finite input-state gain
   - robustness margin wrt unmodeled dynamics
   - robustness margin wrt delayed dynamics
3. periodic input, periodic output
4. modularity and interconnection properties
5. accurate numerical integration and equilibrium point computation



search for contraction properties

design engineering systems to be contracting

verify correct/safe behavior via known Lipschitz constants

Outline

Convexity and contractivity

Kachurovskii's Theorem: For differentiable $f : \mathbb{R}^n \to \mathbb{R}$, equivalent statements:

1. f is **strongly convex** with parameter ν (and minimum x^*)
2. $-\nabla f$ is **ν-strongly infinitesimally contracting** (with equilibrium x^*), that is

$$\big(-\nabla f(x) + \nabla f(y)\big)^\top (x - y) \leq -\nu \|x - y\|_2^2$$

R. I. Kachurovskii. Monotone operators and convex functionals. *Uspekhi Matematicheskikh Nauk*, 15(4):213–215, 1960

Example #1: Gradient dynamics for strongly convex function

Given differentiable, strongly convex $f : \mathbb{R}^n \to \mathbb{R}$ with parameter $\nu > 0$, **gradient dynamics**

$$\dot{x} = \mathsf{F}_{\mathsf{G}}(x) := -\nabla f(x)$$

F_{G} is infinitesimally contracting wrt $\|\cdot\|_2$ with rate ν

unique globally exp stable point is global minimum

Example #2: Primal-dual gradient dynamics

strongly convex function f s.t. $0 \prec \nu_{\mathsf{min}} I_n \preceq \mathrm{Hess}\, f \preceq \nu_{\mathsf{max}} I_n$

constraint matrix A s.t. $0 \prec a_{\mathsf{min}} I_m \preceq AA^\top \preceq a_{\mathsf{max}} I_m$ (independent rows)

linearly constrained optimization:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subj. to} \quad & Ax = b \end{aligned}$$

primal-dual gradient dynamics:

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\mathsf{PDG}}(x, \lambda) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$$

$\mathsf{F}_{\mathsf{PDG}}$ is infinitesimally contracting wrt $\|\cdot\|_{2,P^{1/2}}$ with rate c

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix} \text{ with } \alpha = \frac{1}{2} \min\left\{\frac{1}{\nu_{\mathsf{max}}}, \frac{\nu_{\mathsf{min}}}{a_{\mathsf{max}}}\right\} \quad \text{and} \quad c = \frac{1}{4} \min\left\{\frac{a_{\mathsf{min}}}{\nu_{\mathsf{max}}}, \frac{a_{\mathsf{min}}}{a_{\mathsf{max}}} \nu_{\mathsf{min}}\right\}$$

Example #3: Laplacian-based distributed gradient

Given $\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\top / n$ = orthogonal projection onto $\mathrm{span}\{\mathbb{1}_n\}^\perp$,

$$0 \prec \lambda_2 \Pi_n \preceq L \preceq \lambda_n I_n$$

decomposable cost: $\min_{x \in \mathbb{R}} \sum_{i=1}^n f_i(x)$ where each f_i is ν_i-strongly convex

$$\begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & \sum_{j=1}^n a_{ij}(x_i - x_j) = 0 \end{cases}$$

Laplacian-based distributed gradient (primal-dual gradient, $2n$ vars):

$$\begin{cases} \dot{x}_{[i]} = -\nabla f_i(x_{[i]}) - \sum_{j=1}^n a_{ij}(\lambda_i - \lambda_j) & \text{for each node } i \\ \dot{\lambda}_i = \sum_{j=1}^n a_{ij}(x_i - x_j) & \text{for each node } i \end{cases}$$

$\mathsf{F}_{\text{Laplacian-DistributedG}}$ **is infinitesimally contracting**† **with** $c = \frac{1}{4}\left(\frac{\lambda_2}{\lambda_n}\right)^2 \min_i \nu_i$

composite minimization (cost = sum of terms with structurally different properties):

$$x^\star \;=\; \operatorname*{argmin}_{x \in \mathbb{R}^n} \; f(x,u) \;+\; g(x)$$

$f(x,u)$ is convex and differentiable in x $\qquad$ $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, closed, and proper (ccp)

proximal operator: $\qquad \operatorname{prox}_{\gamma g}(z) \;:=\; \operatorname*{argmin}_{x \in \mathbb{R}^n} g(x) + \frac{1}{2\gamma}\|x - z\|_2^2$

generalized form of projection for nonsmooth/constrained/large-scale/distributed optimization

Detour: Composite optimization and the proximal operator

composite minimization (cost = sum of terms with structurally different properties):

$$x^\star \;=\; \operatorname*{argmin}_{x \in \mathbb{R}^n} \; f(x,u) \;+\; g(x)$$

$f(x,u)$ is convex and differentiable in x

$g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, closed, and proper (ccp)

proximal operator:

$$\operatorname{prox}_{\gamma g}(z) \;:=\; \operatorname*{argmin}_{x \in \mathbb{R}^n} g(x) + \frac{1}{2\gamma}\|x - z\|_2^2$$

generalized form of projection for nonsmooth/constrained/large-scale/distributed optimization

Equivalence property:

1. x^* is minimizer for: $\displaystyle\min_{x \in \mathbb{R}^n} f(x,u) + g(x)$
2. x^* is fixed point for: $x = \operatorname{prox}_{\gamma g}(x - \gamma \nabla f(x,u))$ for all γ

Example #4: Proximal gradient dynamics

Equivalence property motivates:

proximal gradient dynamics:

$$\dot{x} = \mathsf{F}_{\mathsf{ProxG}}(x) := -x + \operatorname{prox}_{\gamma g}(x - \gamma \nabla f(x))$$

projected gradient descent is special case

$\mathsf{F}_{\mathsf{ProxG}}$ is infinitesimally contracting wrt $\|\cdot\|_2$

for $0 < \gamma < \dfrac{2}{\ell}$, with rate $c = 1 - \max\{|1-\gamma\nu|, |1-\gamma\ell|\}$,

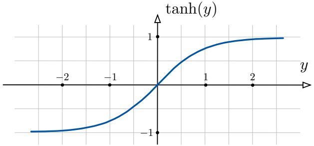
for $\gamma^\star = \dfrac{2}{\nu+\ell}$, with maximal rate $c^\star = \dfrac{2\nu}{\nu+\ell}$

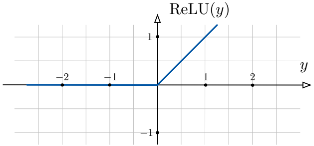
$$\dot{x} = \mathsf{F}_{\mathsf{FR}}(x) := -x + \Phi(Wx + Bu)$$

sigmoid, hyperbolic tangent

ReLU $= \max\{x, 0\} = (x)_+$

$0 \leq \Phi_i'(y) \leq 1$





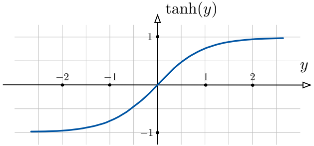
F_{FR} is infinitesimally contracting wrt $\|\cdot\|_\infty$ with rate $1 - \mu_\infty(W)_+$ if

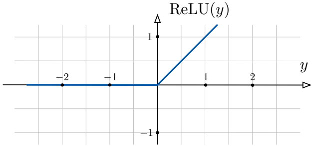
$\mu_\infty(W) < 1$ $\quad$ $\left(\text{i.e., } w_{ii} + \sum_j |w_{ij}| < 1 \text{ for all } i\right)$

Example #5: Firing-rate recurrent neural network

$$\dot{x} = \mathsf{F}_{\mathsf{FR}}(x) := -x + \Phi(Wx + Bu)$$

sigmoid, hyperbolic tangent
ReLU $= \max\{x, 0\} = (x)_+$
$0 \leq \Phi_i'(y) \leq 1$





F_{FR} is infinitesimally contracting wrt $\|\cdot\|_\infty$ with rate $1 - \mu_\infty(W)_+$ if

$\mu_\infty(W) < 1$ $\quad$ $\big($i.e., $w_{ii} + \sum_j |w_{ij}| < 1$ for all $i\big)$

Note: clear **graphical interpretation** + **generalization to interconnection theorem**

Example #6: Firing-rate network with symmetric synapses

$$\dot{x} = \mathsf{F}_{\mathsf{FR}}(x) := -x + \Phi(Wx + Bu)$$
$$0 \le \Phi_i'(y) \le 1 \qquad \text{and} \qquad W = W^\top \text{ with } \lambda_W = \lambda_{\mathsf{max}}(W)$$

For $\lambda_W < 1$ and $\lambda_W \neq 0$, F_{FR} **is infinitesimally contracting with rate** $-1 + (\lambda_W)_+$

For $\lambda_W = 1$, F_{FR} **is weakly infinitesimally contracting**

Note: when $W = W^\top$, **sharper result**, but no graph interpretation and hard to generalize

Assume $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

- $x \mapsto f(x,y)$ is ν_x-strongly convex, uniformly in y
- $y \mapsto f(x,y)$ is ν_y-strongly concave, uniformly in x

saddle dynamics (primal-descent / dual-ascent):

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \mathsf{F}_{\mathsf{S}}(x,y) := \begin{bmatrix}-\nabla_x f(x,y)\\ \nabla_y f(x,y)\end{bmatrix}$$

F_{S} is infinitesimally contracting wrt $\|\cdot\|_2$ with rate $\min\{\nu_x,\nu_y\}$

unique globally exp stable point is saddle point (min in x, max in y)

Example #8: Pseudogradient play

Each player i aims to minimize its own cost function $J_i(x_i, x_{-i})$ (not a potential game)

pseudogradient dynamics (aka gradient play in game theory):

$$\dot{x} = \mathsf{F}_{\mathsf{PseudoG}}(x) = -\big(\nabla_1 J_1(x_1, x_{-1}), \dots, \nabla_n J_n(x_n, x_{-n})\big) \qquad \text{(stacked vector)}$$

$$\iff \qquad \dot{x}_i = -\nabla_i J_i(x_i, x_{-i})$$

- **strong convexity wrt x_i:** J_i is μ_i strongly convex wrt x_i, uniformly in x_{-i}
- **Lipschitz wrt x_{-i}:** $\mathrm{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$, uniformly in x_{-j}
- $\mathsf{F}_{\mathsf{PseudoG}}$ gain matrix is Hurwitz

$\implies$ **$\mathsf{F}_{\mathsf{PseudoG}}$ is infinitesimally contracting wrt appropriate diag-weighted $\|\cdot\|_2$**

Example #9: Best response play

Each player i aims to minimize its own cost function $J_i(x_i, x_{-i})$
$\mathsf{BR}_i : x_{-i} \to \operatorname{argmin}_{x_i} J_i(x_i, x_{-i})$ best response of player i wrt other decisions x_{-i}

best response dynamics:

$$\dot{x} = \mathsf{F}_{\mathsf{BR}}(x) := \mathsf{BR}(x) - x$$
$$\iff \quad \dot{x}_i = \mathsf{BR}_i(x_{-i}) - x_i$$

- **strong convexity wrt x_i:** J_i is μ_i strongly convex wrt x_i, uniformly in x_{-i}
- **Lipschitz wrt x_{-i}:** $\mathsf{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$, uniformly in x_{-j}
  $\implies$ BR_i **is Lipschitz wrt x_j with constant ℓ_{ij}/μ_i**
- F_{BR} gain matrix is Hurwitz $\iff$ BR is a discrete-time contraction
  $\implies$ $\mathsf{BR} - \mathsf{Id}$ **is infinitesimally contracting wrt appropriate diag-weighted** $\|\cdot\|_2$

Equivalent statements:

1. $\text{F}_{\text{PseudoG}}$ gain matrix: $\begin{bmatrix} -\mu_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -\mu_n \end{bmatrix}$ is Hurwitz
2. F_{BR} gain matrix: $\begin{bmatrix} -1 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & -1 \end{bmatrix}$ is Hurwitz
3. discrete-time F_{BR} gain matrix: $\begin{bmatrix} 0 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & 0 \end{bmatrix}$ is Schur

Aggregative games: $J_i(x_i, x_{-i}) = f_i\big(x_i, \frac{1}{n}\sum_{j=1}^n x_j\big)$
assume f_i is μ_i-strongly convex wrt x_i and $\ell_i = \text{Lip}_y(\nabla_{x_i} f_i(x_i, y))$

$$\mu_i > \ell_i \text{ for each agent } i \qquad \Longrightarrow \qquad \text{gain matrix is Hurwitz}$$

Outline

Many convex optimization problems can be solved with contracting dynamics

$$\dot{x} = \mathsf{F}(x)$$

Parametric and time-varying convex optimization

Many convex optimization problems can be solved with contracting dynamics

$$\dot{x} = \mathsf{F}(x)$$

contracting dynamics for parametric strongly-convex optimization

$$\dot{x} = \mathsf{F}(x, \theta)$$

contracting dynamics for time-varying strongly-convex optimization

$$\dot{x} = \mathsf{F}\big(x, \theta(t)\big)$$

Equilibrium tracking

For parameter-dependent vector field $\mathsf{F} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ and differentiable $\theta : \mathbb{R}_{\geq 0} \to \Theta \subset \mathbb{R}^d$

$$\dot{x}(t) = \mathsf{F}(x(t), \theta(t))$$

Assume there exist norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\Theta}$ s.t.

- **contractivity wrt x:** $\mathrm{osLip}_x(\mathsf{F}) \leq -c < 0$, uniformly in θ
- **Lipschitz wrt θ:** $\mathrm{Lip}_\theta(\mathsf{F}) \leq \ell$, uniformly in x

Theorem: Equilibrium tracking for contracting dynamics

1. for each fixed θ, there exists a unique equilbrium $x^\star(\theta)$
2. the equilibrium map $x^\star(\cdot)$ is Lipschitz with constant $\dfrac{\ell}{c}$

Equilibrium tracking

For parameter-dependent vector field $\mathsf{F} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ and differentiable $\theta : \mathbb{R}_{\geq 0} \to \Theta \subset \mathbb{R}^d$

$$\dot{x}(t) = \mathsf{F}(x(t), \theta(t))$$

Assume there exist norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\Theta}$ s.t.

- **contractivity wrt x**: $\mathrm{osLip}_x(\mathsf{F}) \leq -c < 0$, uniformly in θ
- **Lipschitz wrt θ**: $\mathrm{Lip}_\theta(\mathsf{F}) \leq \ell$, uniformly in x

Theorem: Equilibrium tracking for contracting dynamics

1. for each fixed θ, there exists a unique equilbrium $x^\star(\theta)$
2. the equilibrium map $x^\star(\cdot)$ is Lipschitz with constant $\dfrac{\ell}{c}$
3. $D^+\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \dfrac{\ell}{c}\,\|\dot{\theta}(t)\|_{\Theta}$

Consequences for tracking error

$$D^+\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \frac{\ell}{c}\,\|\dot{\theta}(t)\|_{\Theta}$$

Consequences for tracking error

$$D^+\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \frac{\ell}{c}\,\|\dot{\theta}(t)\|_{\Theta}$$

- bounded input, bounded error
  with asymptotic bound:

$$\limsup_{t\to\infty} \|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \quad\leq\quad \frac{\ell}{c^2}\,\limsup_{t\to\infty} \|\dot{\theta}(t)\|_{\Theta}$$

Consequences for tracking error

$$D^+\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \frac{\ell}{c}\,\|\dot\theta(t)\|_\Theta$$

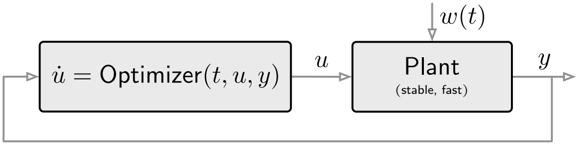
- bounded input, bounded error
  with asymptotic bound:

$$\limsup_{t\to\infty}\|x(t)-x^\star(\theta(t))\|_{\mathcal{X}} \quad\leq\quad \frac{\ell}{c^2}\,\limsup_{t\to\infty}\|\dot\theta(t)\|_\Theta$$

- bounded energy input, bounded energy error
- vanishing input, vanishing error
- exponentially vanishing input $\sim \mathrm{e}^{-ht}$, exponentially vanishing error $\sim \mathrm{e}^{-\min\{c,h\}t}$
- periodic input, periodic error

Outline

Application: Online feedback optimization



online feedback optimization

online optimization, optimization-based feedback, input/output regulation ...

$$\begin{cases} \min & \text{cost}_1(u) + \text{cost}_2(y) \\ \text{subj. to} & y = \text{Plant}\big(u, w(t)\big) \end{cases} \quad \Longrightarrow \quad \begin{cases} \dot{u} = \text{Optimizer}(t, u, y) \\ y = \text{Plant}\big(u, w(t)\big) \end{cases}$$

A. Hauswirth, S. Bolognani, G. Hug, and F. Dorfler. Timescale separation in autonomous optimization. *IEEE Transactions on Automatic Control*, 66(2):611–624, 2021.

G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese. Time-varying optimization of LTI systems via projected primal-dual gradient flows. *IEEE Transactions on Control of Network Systems*, 9(1):474–486, 2022.

Example #10: Gradient controller

- fast/stable LTI plant with control input u and state/measurement disturbance $w(t)$:

$$\begin{aligned} \epsilon \dot{x} &= Ax + Bu + Ew(t) \qquad\qquad A \text{ Hurwitz} \\ y &= Cx + Dw(t) \end{aligned}$$

- in singular perturbation limit as $\epsilon \to 0^+$, **steady state map (Y_u and Y_w)**

$$y = \underbrace{-CA^{-1}B}_{=:\ Y_u} u + \underbrace{(D - CA^{-1}E)}_{=:\ Y_w} w$$

Example #10: Gradient controller

- fast/stable LTI plant with control input u and state/measurement disturbance $w(t)$:

$$\begin{aligned} \epsilon \dot{x} &= Ax + Bu + Ew(t) \qquad\qquad A \text{ Hurwitz} \\ y &= Cx + Dw(t) \end{aligned}$$

- in singular perturbation limit as $\epsilon \to 0^+$, **steady state map (Y_u and Y_w)**

$$y \;=\; \underbrace{-CA^{-1}B}_{=:\,Y_u}\, u \;+\; \underbrace{(D - CA^{-1}E)}_{=:\,Y_w}\, w$$

- define **cost function $\mathcal{E}$ on u and y**:

$$\mathcal{E}(u, w) = \phi(u) + \psi(Y_u u + Y_w w), \qquad (\phi \text{ is } \nu\text{-strongly convex and } \psi \text{ is convex})$$

and note

$$\begin{aligned} \nabla_u \mathcal{E}(u, w) &= \nabla\phi(u) + Y_u^\top \nabla\psi(Y_u u + Y_w w) \\ &= \nabla\phi(u) + Y_u^\top \nabla\psi(y) \qquad\qquad (\text{no need to measure } w(t)) \end{aligned}$$

Example #10: Gradient controller

equilibrium trajectory let $u^*(t)$ be solution to

$$\begin{aligned} \min_{u} \quad & \phi(u) + \psi(y(t)) \\ \text{subj to} \quad & y(t) = Y_u u + Y_w w(t) \end{aligned} \qquad (\nu\text{-strongly convex } \phi, \text{ convex } \psi)$$

gradient controller

$$\dot{u} = \mathsf{F}_{\mathsf{GradCtrl}}(u, w) \;:=\; -\nabla \mathcal{E}_u(u, w) = -\nabla \phi(u) - Y_u^\top \nabla \psi(Y_u u + Y_w w)$$

Example #10: Gradient controller

equilibrium trajectory let $u^*(t)$ be solution to

$$\begin{aligned} \min_u \quad & \phi(u) + \psi(y(t)) \\ \text{subj to} \quad & y(t) = Y_u u + Y_w w(t) \end{aligned}$$

(ν-strongly convex ϕ, convex ψ)

gradient controller

$$\dot{u} = \mathsf{F}_{\mathsf{GradCtrl}}(u,w) \;:=\; -\nabla \mathcal{E}_u(u,w) = -\nabla\phi(u) - Y_u^\top \nabla\psi(Y_u u + Y_w w)$$

Equilibrium tracking for the gradient controller

1. $\mathrm{osLip}_u(\mathsf{F}_{\mathsf{GradCtrl}}) \leq -\nu$ (gradient of ν-strongly convex function)
2. $\mathrm{Lip}_w(\mathsf{F}_{\mathsf{GradCtrl}}) = \ell_w := \|Y_u^\top\| \,\mathrm{Lip}(\nabla\psi) \|Y_w\|$

$$\limsup_{t\to\infty} \|u(t) - u^*(t)\| \quad \leq \quad \frac{\ell_w}{\nu^2} \limsup_{t\to\infty} \|\dot{w}(t)\|$$

Example #11: Projected gradient controller

Constrained feedback optimization:

$$\min_u \quad \mathcal{E}(u,w) = \phi(u) + \psi(Y_u u + Y_w w) \qquad (\nu \text{ strongly convex, } \ell_u \text{ strongly smooth, } \ell_w)$$

$$\text{subj. to} \quad u \in \mathcal{U} \qquad (\text{nonempty, closed, convex. } P_{\mathcal{U}} = \text{orthogonal projection})$$

Projected gradient controller (example of proximal gradient dynamics):

$$\dot{u} = \mathsf{F}_{\mathsf{PGC}}(u,w) \; := -u + \; P_{\mathcal{U}}\big(u - \gamma \nabla_u \mathcal{E}(u,w)\big)$$

Equilibrium tracking for projected gradient controller At $\gamma = \frac{2}{\nu+\ell_u}$,

1. $\text{osLip}_u(\mathsf{F}_{\mathsf{PGC}}) \leq -c_{\mathsf{PGC}} := -\dfrac{2\nu}{\nu+\ell_u}$ (contractivity prox gradient)
2. $\text{Lip}_w(\mathsf{F}_{\mathsf{PGC}}) = \ell_{\mathsf{PGC}} := \dfrac{2}{\nu+\ell_u}\ell_w$

$$\limsup_{t\to\infty} \|u(t) - u^*(t)\| \quad \leq \quad \frac{\ell_{\mathsf{PGC}}}{c_{\mathsf{PGC}}^2} \limsup_{t\to\infty} \|\dot{w}(t)\| \qquad \text{(eq tracking)}$$

Outline

Conclusions

contractivity = robust computationally-friendly stability

fixed point theory + Lyapunov stability theory + geometry of metric spaces

- theory
- examples
- control application

Ongoing work

1. applications to ML and biologically-inspired neural networks
2. applications to optimization-based control designs:
   model predictive control, control barrier functions, low-gain integral control
3. equilibrium tracking with noise
   applications to optimization-based control