Optimization and control: mutual connections

Boris Polyak
boris@ipu.ru

Institute of Control Sciences, RAS, Moscow

Conference “Optimization without borders”
SIRIUS
July, 2021

These are just some thoughts on development of two scientific disciplines...

$\min f(x), \quad x \in Q \subset R^n$

Extensions: Infinite dimensions, Matrix variables, Stochastic problems, Robust optimization...

Optimality conditions: Fermat 1629, Lagrange 1790, Karush, Kuhn-Tucker 1936–1950 , Kantorovich 1940, Dubovitski-Miljutin 1963.

Methods: Newton 1690, Cauchy 1847, Kantorovich 1948, Dantzig 1948, Nemirovski 1976, Nesterov-Nemirovski 1989, 1994.

Role of convexity: Rockafellar 1970, Pshenichny.

Watt regulator, transfer function, input-output description, step response, Laplace transform, stability, Nyquist plot, Routh-Hurwitz test, Lyapunov methods...

Keywords for Control and Optimization were completely different.

Control revolution 1960

Bellman, Kalman, Pontryagin

- State-space formulation $\dot{x} = f(x, u, t)$,
  $x(t)$ - state variable, $u(t)$ - control.
- Performance index Optimal control
  $\min F(x(T))$ or $\min \int_0^T \phi(x, u, t)dt$ or $\min T$

First rendezvous: optimal control as optimization in Hilbert space

$$J(u) = F(x(T)), \dot{x} = f(x, u, t), u \in L_2(0, T)$$

Gradient method (or projected gradient)

$$u_{k+1} = u_k - \gamma_k \nabla J(u_k)$$

Conjugate equation for calculation of the gradient.

Kelley 1960, Bryson-Denham 1961, Polak; Demyanov, Pshenichny,

Levitin-Polyak 1966

Maximum principle as optimality condition.

Dubovitski-Miljutin, Girsanov

$u(x,t)$ as variable. Feedback vs open-loop control. Bellman function

$$V(x,t) = \min \int_t^T \phi(x(\tau), u, \tau) d\tau, x(t) = x$$

Jacobi-Bellman equation.

Discrete-time optimal control = nonlinear programming.

Tabak-Kuo, Optimal control by mathematical programming, 1971;

calculations: Fedorenko, Mikhalevich-Shor, Krylov-Chernousko, Moiseev.

Linear Quadratic Regulator

Kalman 1960

$$\dot{x} = Ax + Bu, \quad x(0), x(T),$$

$$\min \int_0^T ((Qx, x) + (Ru, u))dt$$

$V(x, t)$ is quadratic, explicit solution via two-point ODE

LQR, infinite horizon

Kalman 1960

$T = \infty$. Explicit solution. Optimal $u(t)$ can be written as $u = -Kx$ – static state feedback. Finding K via Riccati equation:

$$K = -R^{-1}B^T X$$

$X \succ 0$ is the solution of ARE

$$A^T X + XA - XBR^{-1}B^T X + Q = 0$$

Feedback optimization

$u = -Kx$, K is independent variable

$$\dot{x}(t) = A_K x(t), \quad A_K = A - BK, Ex(0)x(0)^T = \Sigma$$

Matrix optimization problem

$$f(K) = \mathrm{Tr}\left(X\Sigma\right) \to \min_{K \in S}$$

$X \succ 0$ is the solution of Lyapunov equation

$$A_K^\top X + XA_K + K^\top RK + Q = 0,$$

Gradient method

$$K_{j+1} = K_j - \gamma_j \nabla f(K_j), K_j \in S = \{K : A_K \text{ is stable}\}$$

Formula for $\nabla f(K_j)$ — Kalman 1960, Levine-Athans 1970 (solution of additional Lyapunov equation). But S is nonconvex, $f(K)$ is nonconvex and coercive.

Validation: Bu-Mesbahi-Fazel-Mesbahi 2019; Mohammadi-Zare-Soltanolkotabi-Jovanovic 2019; Fatkhullin-Polyak 2020.

Main results: $f(K)$ is L-smooth and gradient-dominated on S_0; gradient method converges.

Extensions

- Output feedback $u = Ky, y = Cx$. Gradient method converges to stationary point.
- Problems with external disturbances etc.
- Parametric control (e.g. PID) $u = \sum_{i=1}^{m} k_i A_i x$.

Challenges for optimization

- $f(K)$ is unbounded on S. How to design a nice first order method for minimization? We can use also second directional derivative.
- Stability domain S is bad, however feasibility oracle is simple. How to adjust step-size rule?
- When gradient-domination condition holds?
- Extension of LMIs:

$$A(k) = A_0 + \sum_{i=1}^{m} k_i A_i, \quad S = \{k : A(k) \text{ is stable}\}$$

S is convex if A_i commute.

LMI breakthrough

Nesterov-Nemirovski 1990

LMI $S : A(x) = A_0 + \sum_{i=1}^{n} x_i A_i \succ 0, \quad A_i = A_i^T$

SDP $\min f(x), x \in S, f(x)$ convex.

Self-concordant barrier

Interior-point method

Applications to convex optimization, LP, discrete optimization (convex relaxation) etc. CVX and other software.

LMI in control

Boyd 1994

Stability analysis

A Hurwitz $\leftrightarrow \exists P \succ 0, AP + PA^T \prec 0$

Synthesis

$\dot{x} = Ax + Bu,\ u = Kx$ is stable $\leftrightarrow \exists P \succ 0$

$AP + PA^T + BY + Y^T B^T \prec 0, K = YP^{-1}$

Thus main control problems (analysis, stabilization, design, performance) can be reduced to LMI and SDP. Now LMI technique is the mainstream in control.

Stability: nonlinear systems

Lyapunov 1892

Simplest case: $\dot{x} = \phi(x), x(0)$.
Lyapunov function $V(x) > 0, V(0) = 0$, if $\dot{V}(x) = (\nabla V(x), \phi(x)) < 0$, then $x(t) \longrightarrow 0$ (absolute stability).
Linear system: $\phi(x) = Ax, V(x) = (Px, x), \dot{V}(x) = ((PA + A^T P)x, x)$. Stability: $P \succ 0, PA + A^T P \prec 0$.

Stability analysis vs optimization algorithms

Optimization: $\min f(x)$; gradient flow:

$\dot{x} = -\nabla f(x)$

Asymptotic Stability:

$\dot{x} = \phi(x)$; does $x(t) \longrightarrow x^*$?

	Optimization	Stability
Data	Function	Equation
	Design a method	Find $V(x)$
Goal	Prove convergence	Prove stability

Modified Lyapunov theorem

Theorem Assume $V(x) \geq v^*$, $-\dot{V}(x) \geq W(x) \geq 0$ for all x, $\dot{x}(t) = \phi(x)$. Then

1. $\liminf W\big(x(t)\big) = 0$ as $t \to \infty$ and

$$\min_{0 \leq t \leq T} W\big(x(t)\big) \leq \frac{V\big(x(0)\big) - v^*}{T}.$$

2. If $W(x)$ is convex, $\bar{x}(T) = \frac{1}{T}\int_0^T x(t)\mathrm{d}t$, then

$$W\big(\bar{x}(T)\big) \leq \frac{V\big(x(0)\big) - v^*}{T}.$$

3. If (i) $V(x^*) = v^*$ with $V(x) > v^*$ for $x \neq x^*$, (ii) $\inf_{||x-x^*|| \geq R}\big(V(x) - v^*\big) \geq \varepsilon > 0$ for $R > 0$, and (iii) $W(x^*) = 0$ with $W(x) > 0$ for $x \neq x^*$, Then $x(t) \to x^*$ as $t \to \infty$.
4. If $W(x) \geq \gamma\big(V(x) - v^*\big)$, $\gamma > 0$, then

$$V\big(x(t)\big) - v^* \leq \Big(V\big(x(0)\big) - v^*\Big)\exp(-\gamma t).$$

Application to gradient method 1

Cauchy 1847

$\min f(x), \quad x \in R^n,\ f(x) \in C^1, \dot{x} = -\nabla f(x).$

Let $f(x) \geq f^*$. Then

$$\min_{0 \leq t \leq T} \|\nabla f(x(t)\|^2 \leq \frac{f\big(x(0)\big) - f^*}{T}.$$

If $f(x) - f^* \leq 2\ell \|\nabla f(x)\|^2, \ell > 0$, (PL-cond.), then the convergence $f(x(t)) \to f^*$ is exponential.

Hint: $V(x) = f(x) - f^* \Longrightarrow W(x) = \|\nabla f(x)\|^2$.

Use 1 and 4 of Theorem. Convexity of $f(x)$ and uniqueness of x^* are not required!

Gradient method 2

If $f(x)$ is convex and there exists x^* such that $f(x^*) = f^*$, then

$$f(\bar{x}(T)) - f^* \leq \frac{\|x(0) - x^*\|^2}{2T}, \bar{x}(T) = \frac{\int_0^T x(t)\mathrm{d}t}{T}.$$

If $f(x)$ is strongly convex, then $x(t) \to x^*$ exponentially as $t \to \infty$

Hint: $V(x) = ||x - x^*||^2 \Longrightarrow W(x) = f(x) - f^*$.

Use 2 and 4.

Heavy ball

Polyak 1964

$\min f(x), \quad x \in R^n$

$\ddot{x} + a\dot{x} + b\nabla f(x) = 0, a > 0, b > 0.$

Global convergence rate for quadratic function, local - for nonquadratic. Proof:

$$\dot{x} = y,$$
$$\dot{y} = -ay - b\nabla f(x).$$

Full energy as Lyapunov function:

$V = f(x) + \frac{1}{2b}\|y\|^2$, then $W = \frac{a}{b}\|y\|^2 \geq 0$, can apply La Salle invariance principle to validate convergence, but not convergence rate.

New Lyapunov functions

Polyak-Scherbakov, IJC, 2018

Let $f(x) \geq f^*, f \in C^2, ||\nabla^2 f(x)|| \leq L$,

$$V = f(x) - f^* + \frac{a}{a^2 + 2bL}(\nabla f(x), y) + \frac{L\|y\|^2}{a^2 + 2bL}.$$

Then for any $x(0), y(0)$, for the HB

$$\min_{0 \leq t \leq T} ||\nabla f(x(t))||^2 \leq \frac{V(0)}{cT}, \quad c = \frac{ab}{a^2 + 2bL}.$$

If PL condition holds, then x^* exists, and the convergence $x(t) \to x^*$ is exponential.

Extensions

- Convex case.
- For nonstationary damping $a(t)$ fast rate of convergence $O(1/T^2)$ can be obtained Su-Boyd-Candes 2014, Attouch, Bolte ....
- Discrete-time versions.
- Use of control tools (such as LQC) to design other accelerated methods Taylor-VanScoy-Lessard....

General tendency: control+optimization

Examples:

Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives (Muehlebach-Jordan 2021)
A dynamical view on optimization algorithms... (Bu-Xu-Chen 2021)
A Lyapunov analysis of accelerated methods in optimization (Wilson-Recht-Jordan 2021)

+many others

New trends: optimization+control+machine learning

New keywords: policy optimization, adaptive control, reinforcement learning, model-free control etc.

Examples — IFAC 2020 Plenaries:

B.Recht Reflections on learning to control renaissance

Lee Reinforcement learning for process control and beyond

A lot of conferences, workshops etc.

Number of points where optimization meets control are growing!