Prediction-Correction Methods for Time-Varying Convex Optimization

Andrea Simonetto[†], Aryan Mokhtari[⋆], Alec Koppel[⋆],
Geert Leus[†], Alejandro Ribeiro[⋆]

[†]Delft University of Technology, Delft, The Netherlands

[⋆]University of Pennsylvania, Philadelphia, PA

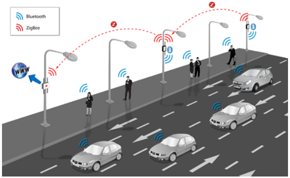
Asilomar Conference on Signals, Systems, and Computers

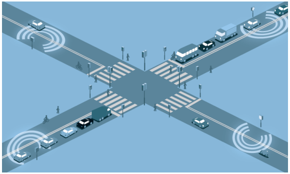
Pacific Grove, CA, Nov. 10, 2015

Continuous-time Optimization

$$\mathbf{x}^*(t) := \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}; t), \quad \text{for } t \geq 0 \, .$$

- Objective $f(\cdot, t)$ $\Rightarrow$ time-varying cost
- For fixed time t $\Rightarrow$ convex program
- Examples problems:
  - $\Rightarrow$ finding control gains
  - $\Rightarrow$ target tracking
  - $\Rightarrow$ statistical model parameters
- Focus: iterative algorithms





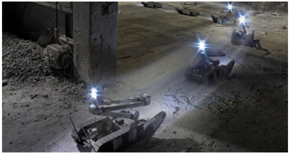


Photo credits: Libelium and iRobot

$$\mathbf{x}^*(t) := \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} f(\mathbf{x}; t), \quad \text{for } t \geq 0 \ .$$

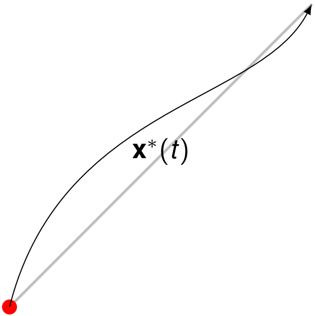
- One approach: sample at discrete times t_k

$$\mathbf{x}^*(t_k) := \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} f(\mathbf{x}; t_k).$$

- Find $\mathbf{x}^*(t_k)$ via Newton or gradient?
  - $\Rightarrow$ only viable if $\mathbf{x}^*(t) \approx \mathbf{x}^*$
  - $\Rightarrow$ otherwise, minimum drifts:

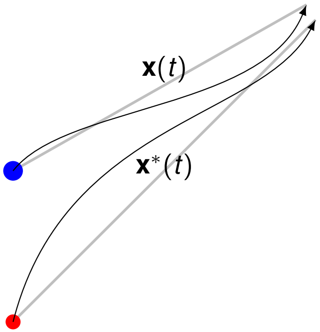
$$\mathbf{x}^*(t_k) \neq \mathbf{x}^*(t_{k+1})$$

- Solution is an entire trajectory.



Prediction-Correction Methods

- Goal: $\mathbf{x}_k = \mathbf{x}(t_k)$ via $f(\mathbf{x}, t_k)$ s.t. $\mathbf{x}_k \to \mathbf{x}^*(t_k)$
  - $\Rightarrow$ Main challenge: distinct $\mathbf{x}^*(t)$ for each t
  - $\Rightarrow$ Stationary convex opt. methods not viable
- Instead, adapt iteration to time variation
  - $\Rightarrow$ Predict $\tilde{\mathbf{x}}(t_{k+1})$ from $f(\cdot, t_k)$
  - $\Rightarrow$ Correct $\tilde{\mathbf{x}}(t_{k+1})$ based on $f(\cdot, t_{k+1})$
- Main idea: track how $\mathbf{x}^*(t_k)$ varies in time



Accurate Solution Tracking is Possible

- Sample the problem at time instances t_k, where $h = t_k - t_{k-1}$
- We develop a method which converges to $O(h^2)$ nbhd. of $\mathbf{x}^*(t)$
  $\Rightarrow$ based on prediction-correction scheme
  $\Rightarrow$ even tighter $O(h^4)$ for Newton-based correction
- Track optimal trajectory $\approx$ error-free

Background

- Non-stationary opt. [Gupal, Kheisin, Nurminskii, etc., 1970s]
  $\Rightarrow$ prediction only, require knowledge of initial optimizer $\mathbf{x}^*(t_0)$
- Parametric programming [Robinson, Rockafellar et.al., 1980s-]
  $\Rightarrow$ prediction only; also require initial optimizer $\mathbf{x}^*(t_0)$
- Distributed signal processing & control (2010-on)
  $\Rightarrow$ Dual decomp. [Jakubeic], ADMM [Boyd], DGD [Nedich]
  $\Rightarrow$ Correction only, arbitrary initialization $\mathbf{x}_0$
- All track $\mathbf{x}^*(t)$ to $O(h)$ neighborhood; sensitive to $\mathbf{x}^*(t_0)$,

Prediction Step via Euler Integration

- Denote $\mathbf{x}_{k+1|k}$ as *prediction* of $\mathbf{x}^*(t_{k+1})$ using objective $f(\mathbf{x}, t_k)$
- Prediction via Newton step

$$\mathbf{x}_{k+1|k} = \mathbf{x}_k - h\left(\nabla_{\mathbf{xx}} f(\mathbf{x}_k; t_k)\right)^{-1} \nabla_{t\mathbf{x}} f(\mathbf{x}_k; t_k)$$

$\Rightarrow$ $\nabla_{\mathbf{xx}} f(\mathbf{x}_k; t_k)$ $\Rightarrow$ Hessian of objective at time t_k

$\Rightarrow$ Step-size h given as sampling rate

$\Rightarrow$ derived via Euler integration of optimality condition residual

Prediction via Residual Tracking

- Prediction step $\Rightarrow$ from optimality condition $\nabla_{\mathbf{x}} f(\mathbf{x}^*(t); t) = \mathbf{0}$
- For any other vector $\mathbf{x}$, we have $\nabla_{\mathbf{x}} f(\mathbf{x}; t) = \mathbf{r}(t)$
  $\Rightarrow$ $\mathbf{r}(t)$ is residual vector

- Track dynamics of residual via first-order difference eqn.

$$\nabla_{\mathbf{x}} f(\mathbf{x}; t) + \nabla_{\mathbf{xx}} f(\mathbf{x}; t)\delta\mathbf{x} + \nabla_{t\mathbf{x}} f(\mathbf{x}; t)\delta t = \mathbf{r}(t),$$

- δt is variation of t. Divide above eqn. by δt, take limit $\delta t \to 0 \ldots$

$$\dot{\mathbf{x}} = -\left(\nabla_{\mathbf{xx}} f(\mathbf{x}; t)\right)^{-1} \nabla_{t\mathbf{x}} f(\mathbf{x}; t)$$

- Using first-order Forward Euler integration, obtain prediction step

Correcting the prediction

- Prediction $\mathbf{x}_{k+1|k}$ of $\mathbf{x}^*(t_{k+1})$ obtained from Euler integration

$$\mathbf{x}_{k+1|k} = \mathbf{x}_k - h\left(\nabla_{\mathbf{xx}} f(\mathbf{x}_k; t_k)\right)^{-1} \nabla_{t\mathbf{x}} f(\mathbf{x}_k; t_k)$$

$\Rightarrow$ uses $f(\mathbf{x}; t_k)$ objective at time t_k

- Correction $\Rightarrow$ either gradient or Newton step using $f(\mathbf{x}; t_{k+1})$

$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1|k} - \begin{cases} \gamma \nabla_{\mathbf{x}} f(\mathbf{x}_{k+1|k}; t_{k+1}) \\ \left(\nabla_{\mathbf{xx}} f(\mathbf{x}_{k+1|k}; t_{k+1})\right)^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_{k+1|k}; t_{k+1}) \end{cases}$$

$\Rightarrow$ Multiple correction steps τ possible for large enough h

GTT: Gradient Trajectory Tracking

Penn
UNIVERSITY of PENNSYLVANIA

- Initialize $\mathbf{x}_0$ arbitrarily, set $h = t_{k+1} - t_k$, step-size $\gamma > 0$.
- For $k = 1, 2, \ldots$

1. *Predict* via Euler-integration of sub-optimality residual

$$\mathbf{x}_{k+1|k} = \mathbf{x}_k - h\left(\nabla_{\mathbf{xx}} f(\mathbf{x}_k; t_k)\right)^{-1} \nabla_{t\mathbf{x}} f(\mathbf{x}_k; t_k)$$

$\Rightarrow$ uses $f(\mathbf{x}; t_k)$ objective at time t_k

2. *Correct* predicted trajectory using gradient step

$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1|k} - \gamma \nabla_{\mathbf{x}} f(\mathbf{x}_{k+1|k}; t_{k+1})$$

$\Rightarrow$ uses objective $f(\mathbf{x}; t_{k+1})$ at time t_{k+1}

NTT: Newton Trajectory Tracking

- Initialize $\mathbf{x}_0$ arbitrarily, set $h = t_{k+1} - t_k$, step-size $\gamma > 0$.
- For $k = 1, 2, \ldots$

1. *Predict* via Euler-integration of sub-optimality residual

$$\mathbf{x}_{k+1|k} = \mathbf{x}_k - h\left(\nabla_{\mathbf{xx}} f(\mathbf{x}_k; t_k)\right)^{-1} \nabla_{t\mathbf{x}} f(\mathbf{x}_k; t_k)$$

$\Rightarrow$ uses $f(\mathbf{x}; t_k)$ objective at time t_k

2. *Correct* predicted trajectory using Newton step

$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1|k} - \left(\nabla_{\mathbf{xx}} f(\mathbf{x}_{k+1|k}; t_{k+1})\right)^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_{k+1|k}; t_{k+1})$$

$\Rightarrow$ uses objective $f(\mathbf{x}; t_{k+1})$ at time t_{k+1}

Technical Conditions

- The function f is smooth in $\mathbf{x}$ and uniformly in t:

  $\Rightarrow$ Twice differentiable

  $\Rightarrow$ m-strongly convex $\Rightarrow$ eigenvalues of Hessian lower bounded

  $\Rightarrow$ Gradients of f are L-Lipschitz continuous

- Bounded higher-order derivatives with respect to both $\mathbf{x}$ and t

  $\Rightarrow$ the mapping $t \mapsto \mathbf{x}^*(t)$ is one-to-one and locally Lipschitz

Convergence of GTT

Theorem

i) *The GTT sequence* $\{\mathbf{x}_k\}$ *converges to nbhd. of optimal* $\mathbf{x}^*(t)$

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}^*(t_k)\| = O(h)$$

ii) *If sampling rate h and step-size* γ *are sufficiently small,*

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}^*(t_k)\| = O(h^2)$$

$\Rightarrow$ Convergence rate is *Q*-linear if step-size satisfies $\gamma < 2/L$.

$\Rightarrow$ Sampling rate threshold: $h < \left[\frac{C_0 C_1}{m^2} + \frac{C_2}{m}\right]^{-1} \left(\rho^{-1} - 1\right)$.

$\Rightarrow \rho = \max\{|1-\gamma m|, |1-\gamma L|\} \Rightarrow$ step-size, cvx, Lipschitz constant

Theorem

i) *The NTT sequence* $\{\mathbf{x}_k\}$ *converges to nbhd. of optimal* $\mathbf{x}^*(t)$

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}^*(t_k)\| = O(h^4)$$

ii) *Provided that for any constant* $c > 0$*, sampling rate* h *satisfies*

$$h \leq \min\left\{1, \sqrt{\frac{Qc}{((1+\delta_1)c+\delta_2)^2}}\right\},$$

iii) *Initialization* $\mathbf{x}_0$ *satisfies* $\|\mathbf{x}_0 - \mathbf{x}^*(t_0)\| \leq ch^2$

$\Rightarrow$ Problem constants δ_1, δ_2, and Q defined as

$$\delta_1 := \frac{C_0 C_1}{m^2} + \frac{C_2}{m},\ \delta_2 := \frac{C_0^2 C_1}{2m^3} + \frac{C_0 C_2}{m^2} + \frac{C_3}{2m},\ Q := \frac{2m}{C_1}.$$

Time-Derivative Approximation

- In many application domains, $\nabla_{t\mathbf{x}} f(\mathbf{x}_k; t_k)$ is not known
  $\Rightarrow$ can be approximated by a first-order backward derivative

$$\tilde{\nabla}_{t\mathbf{x}} f(\mathbf{x}_k; t_k) \approx \frac{\nabla_{\mathbf{x}} f(\mathbf{x}_k; t_k) - \nabla_{\mathbf{x}} f(\mathbf{x}_{k-1}; t_{k-1})}{h}$$

  $\Rightarrow$ situations in which exact motion of target unknown

- Can establish comparable convergence guarentees to exact algs.

Approximate Gradient Tracking

Theorem

GTT method with approx. time-derivative in prediction step yields

i) *If step-size satisfies $\gamma < 2/L$, $\{\mathbf{x}_k\} \to \mathbf{x}^*(t_k)$ Q-linearly rate to $O(h)$ bounded error*

$$\|\mathbf{x}_k - \mathbf{x}^*(t_k)\| \leq \rho^k \|\mathbf{x}_0 - \mathbf{x}^*(t_0)\| + \rho\left[h\left[\frac{2C_0}{m}\right] + \frac{h^2}{2}\left[\frac{C_0^2 C_1}{m^3} + \frac{2C_0 C_2}{m^2} + \frac{2C_3}{m}\right]\right]\left[\frac{1-\rho^k}{1-\rho}\right].$$

ii) *For small enough h and $\gamma > 0$, $\{\mathbf{x}_k\} \to \mathbf{x}^*(t_k)$ Q-linearly up to a bounded $O(h^2)$ error*

$$\|\mathbf{x}_k - \mathbf{x}^*(t_k)\| \leq (\rho\sigma)^k \|\mathbf{x}_0 - \mathbf{x}^*(t_0)\| + \rho\,\frac{h^2}{2}\left[\frac{C_0^2 C_1}{m^3} + \frac{2C_0 C_2}{m^2} + \frac{2C_3}{m}\right]\left[\frac{1-(\rho\sigma)^k}{1-\rho\sigma}\right].$$

Theorem

NTT method with approx. time-derivative in prediction step achieves

$$\|\mathbf{x}_k - \mathbf{x}^*(t_k)\| \leq Q^{-1}(\sigma c + \delta_2')^2 h^4 = O(h^4)$$

whenever for any constant $c > 0$*, the sampling increment h satisfies*

$$h \leq \min\left\{1, \sqrt{\left[\frac{Qc}{((1+\delta_1)c + \delta_2')^2}\right]}\right\},$$

and initial error $\|\mathbf{x}_0 - \mathbf{x}^*(t_0)\|$ *satisfies* $\|\mathbf{x}_0 - \mathbf{x}^*(t_0)\| \leq ch^2$ *.*

Target Tracking Example

- Scalar numerical example

$$f(x;t) := \frac{1}{2}(x - \cos(\omega t))^2 + \frac{\kappa}{2}\sin^2(\omega t)\exp(\mu x^2)$$

- Predictor-corrector methods track $\mathbf{x}^*(t)$

  $\Rightarrow$ magnitudes less in tracking error

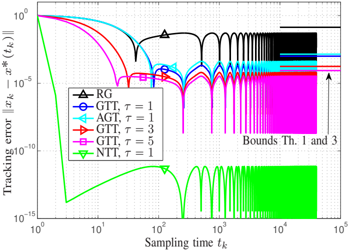
- Worst-case error:

$$\max_{k>\bar{k}}\{\|x_k - x^*(t_k)\|\}$$

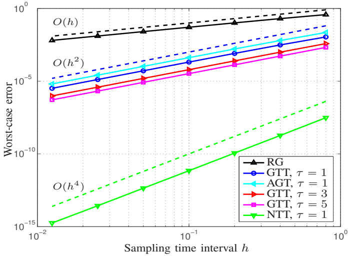
where $\bar{k} = 10^4$.

  $\Rightarrow$ performance improvement clearer

- Newton more accurate but higher cost



(a) Tracked sample planar trajectory



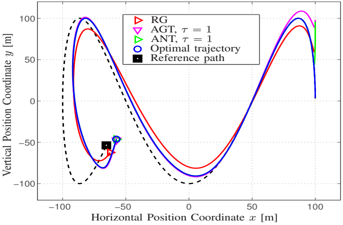
(b) Worst-case error

Target Tracking Example

- Target tracking example

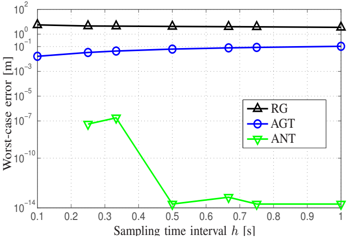
$$f(\mathbf{x}; t) := \frac{1}{2}\left(\|\mathbf{x} - \mathbf{y}(t)\|^2 + \mu_1 \exp(\mu_2 \|\mathbf{x} - \mathbf{b}\|^2)\right)$$

- Track a moving planar object $\mathbf{y}(t)$
  $\Rightarrow$ stay close to base station $\mathbf{b}$
- Predictor-corrector methods track optimal



(a) Tracked sample planar trajectory

- Worst-case error: $\max_{k > \bar{k}} \{\|x_k - x^*(t_k)\|\}$, with $\bar{k} = 10^4$
- Fix comp. cost for given sampling rate
- Newton computationally heavier
  $\Rightarrow$ higher control latency
- Main takeaway: use NTT (ANT) if sampling rate large enough



(b) Worst-case error

Concluding Remarks

- Focused on continuously-varying convex opt. problems
- Brute force approach intractable $\Rightarrow$ discrete-time sampling

- Developed prediction-correction methods

  $\Rightarrow$ Predict where optimal trajectory will be at next time

  $\Rightarrow$ Once we observe info. at next sample time, make correction
- Correction step $\Rightarrow$ Gradient or Newton steps (GTT or NTT)

- Theorem: GTT & NTT track $\mathbf{x}^*(t)$ up to $O(h^2)$ or better

  $\Rightarrow$ Correction reduces tracking error relative to existing methods
- No problem if we don't know exact objective time variation

- Numerical examples illustrate utility in control domain

  $\Rightarrow$ When sampling rate is large enough, better to use NTT

References

- A. Simonetto, A. Koppel, A. Mokhtari, G. Leeus, and A. Ribeiro, “A Class of Prediction-Correction Methods for Time-Varying Convex Optimization,” IEEE Trans. Signal Process (submitted)., Sept. 2015.
- A. Koppel, A. Simonetto, A. Mokhtari, G. Leus, and A. Ribeiro, “Target Tracking with Dynamic Convex Optimization”, IEEE Global Conference on Signal and Information Processing (to appear), Orlando, FL, Dec. 14-16, 2015.
- A. Simonetto, A. Koppel, A. Mokhtari, G. Leeus, and A. Ribeiro “Prediction-Correction Methods for Time-Varying Convex Optimization.” in Proc. Asilomar Conf. on Signals Systems Computers, Pacific Grove, CA, November 8-11 2015.

http://seas.upenn.edu/~akoppel/