The Complex Barycenter Method for Direct Optimization

Felipe Pait
Universidade de São Paulo pait@usp.br

A randomized version of the recently developed barycenter method for derivative-free optimization has desirable properties of a gradient search. We develop a complex version to avoid evaluations at high-gradient points. The method is parallelizable in a natural way and robust under noisy measurements, and has applications to control design.

ML, NN, AI,
Applied Functional
Approximation?

Problem of interest.

$$\min_{x \in \mathcal{X}} f(x)$$

No explicit, differentiable expression for $f(\cdot)$ is available.

Values $f(x_i)$ for a set of x_i obtained by experiment, simulation.

Barycenter — batch and recursive formulas.

Batch:

$$\widehat{x}_n = \frac{\sum_{i=1}^n x_i \mathrm{e}^{-\nu f(x_i)}}{\sum_{i=1}^n \mathrm{e}^{-\nu f(x_i)}}. \tag{1}$$

Recursive:

$$m_n = m_{n-1} + \mathrm{e}^{-\nu f(x_n)}, \tag{2}$$

$$\widehat{x}_n = \frac{1}{m_n}\left(m_{n-1}\widehat{x}_{n-1} + \mathrm{e}^{-\nu f(x_n)} x_n\right). \tag{3}$$

Pick points x_n as sum of barycenter $\widehat{x}_{n-1}$ and "curiosity" z_n:

$$x_n = \widehat{x}_{n-1} + z_n, \tag{4}$$

$$\Delta\widehat{x}_n = \widehat{x}_n - \widehat{x}_{n-1} = \frac{\mathrm{e}^{-\nu f(x_n)}}{m_{n-1} + \mathrm{e}^{-\nu f(x_n)}} z_n. \tag{5}$$

Statement of results.

Theorem 1. *If z_n has a Gaussian distribution, the expected value of $\Delta\widehat{x}_n = \widehat{x}_n - \widehat{x}_{n-1}$ is proportional to negative of the average value of the gradient of $f(\widehat{x}_{n-1} + z_n)$ in the support of the distribution of z.*

Define $F_n(z) = \frac{e^{-\nu f(\widehat{x}_{n-1}+z)}}{m_{n-1}+e^{-\nu f(\widehat{x}_{n-1}+z)}}$, $\bar{F}_n(z) = \frac{m_{n-1}}{m_{n-1}+e^{-\nu f(\widehat{x}_{n-1}+z)}} F_n$.

Theorem 2. *Under the conditions of Theorem 1 and assuming that the variance of z is small, the variance of $\Delta\widehat{x}_n$ for $\bar{z} = 0$ near a critical point of $f(x)$ where $\nabla f = 0$ is approximately*

$$\mathrm{Var}(\Delta\widehat{x}) \approx \Sigma E[F^2] - 2\nu\Sigma^T E\left[F\bar{F}\nabla^2 f\right]\Sigma. \tag{6}$$

Complex coefficient ν.

$$\eta_n^\alpha = \frac{\sum_{i=1}^n x_i^\alpha \mathrm{e}^{-\nu f(x_i)}}{\sum_{i=1}^n \mathrm{e}^{-\nu f(x_i)}}, \tag{7}$$

Now estimate of extremum is (for each coordinate α)

$$\widehat{x}_n^\alpha = |\eta_n^\alpha|. \tag{8}$$

Theorem 3. *The expected contribution of measurements made outside of any region where $\nabla f \approx 0$ is discounted by one factor, proportional to ∇f and to the ratio between the complex magnitude of ν and its real part, for each dimension of the search space.*

Destructive interference motivates employing complex values.

Noise or experimental errors.

Minimize f using noisy oracle answers $f(x_i) + w_i$.

Nominal values $\bar{m} = \sum_{i=1}^{n} \mathrm{e}^{-\nu f(x_i)}$ and $\bar{\eta} = \sum_{i=1}^{n} x_i \mathrm{e}^{-\nu f(x_i)} / \bar{m}$,
scalar quantity $\bar{\bar{m}} = \sum_{i=1}^{n} \mathrm{e}^{-2\nu f(x_i)}$, vector $\bar{\bar{\eta}} = \sum_{i=1}^{n} x_i \mathrm{e}^{-2\nu f(x_i)} / \bar{\bar{m}}$,
matrix $\breve{\eta} = \sum_{i=1}^{n} x_i x_i^T \mathrm{e}^{-2\nu f(x_i)} / \bar{\bar{m}}$.

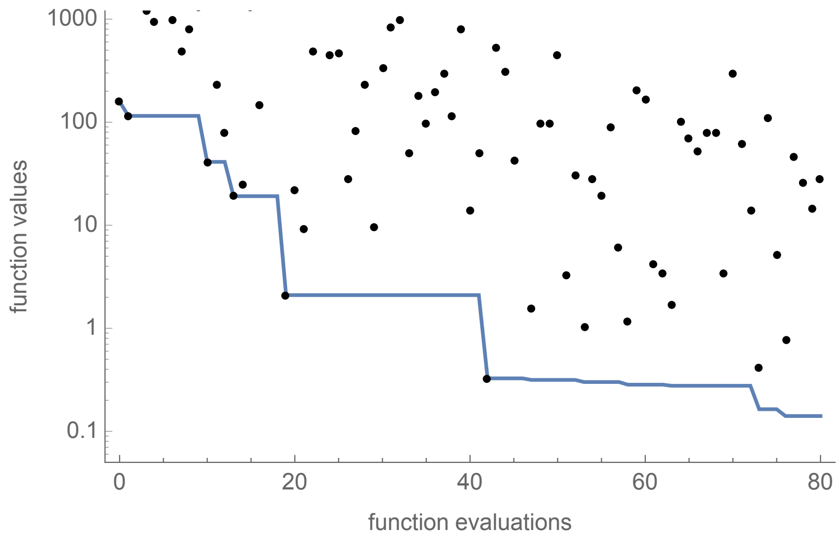
Theorem 4. *Assuming σ^2 (variance of w_i) small, the mean and variance of η are approximately:*

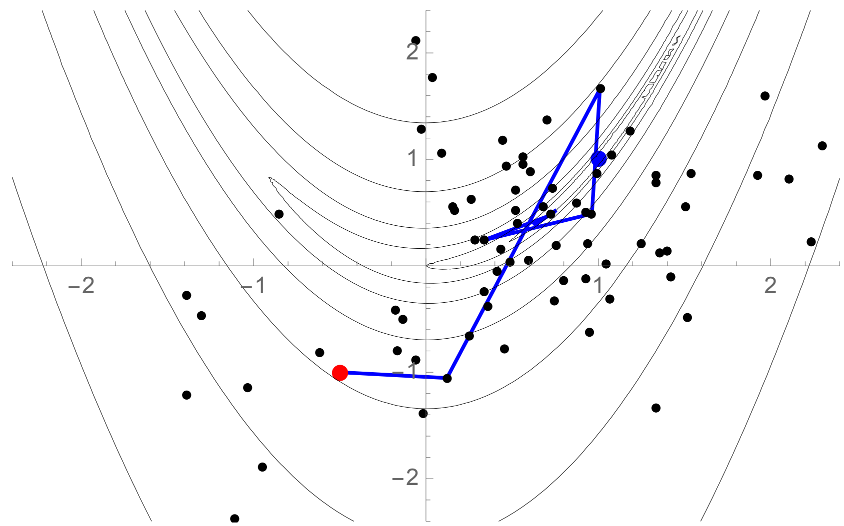
$$E[\eta] \approx \bar{\eta} + \frac{\bar{\bar{m}}}{\bar{m}^2}(\bar{\eta} - \bar{\bar{\eta}})\nu^2\sigma^2 \text{ and} \tag{9}$$

$$\mathrm{Var}[\eta] \approx \frac{\bar{\bar{m}}}{\bar{m}^2}(\bar{\eta}\bar{\eta}^T - \bar{\eta}\bar{\bar{\eta}} - \bar{\bar{\eta}}\bar{\eta} + \breve{\eta})\nu^2\sigma^2. \tag{10}$$

Illustration: Rosenbrock banana function.

$$f(x,y) = 100\left(x^2 - y\right)^2 + (1-x)^2$$

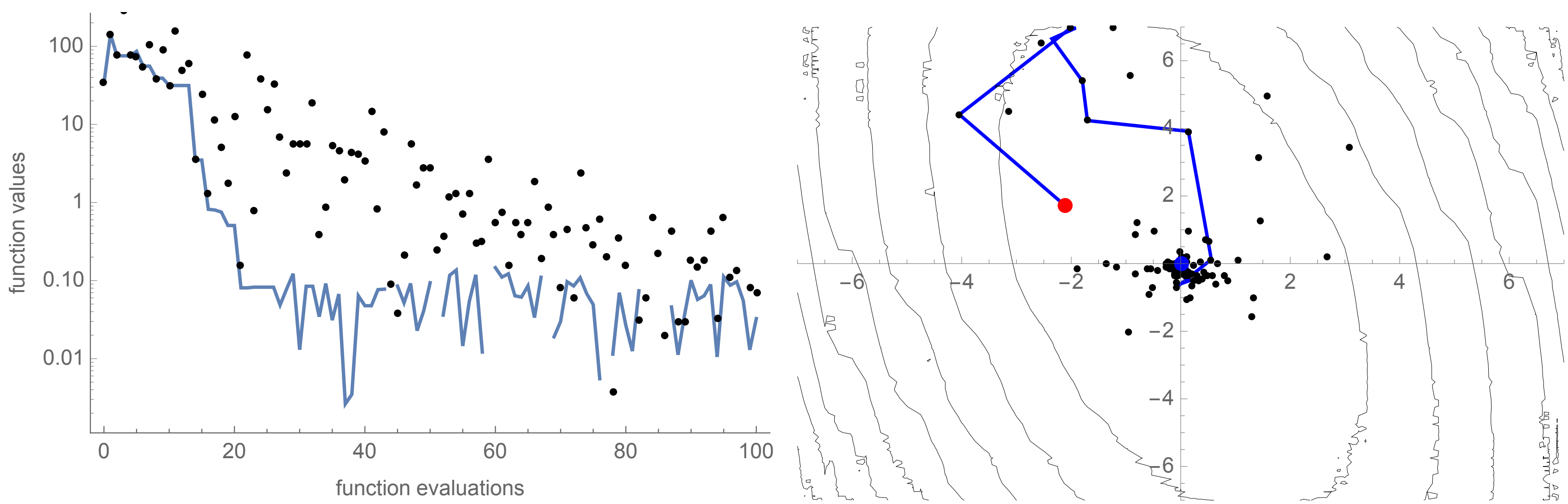




Dots for test points x, line for estimates $\hat{x}$.

Perturbed quadratic function.

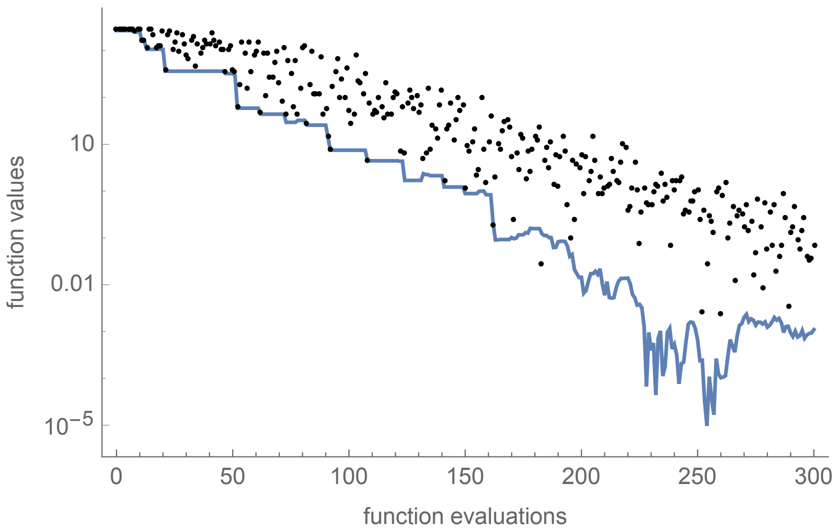
$$f(x,y) = 10x^2\left(1 + \frac{75}{100}\frac{\cos(70x)}{12}\right) + \frac{\cos(100x)^2}{24} + 2y^2\left(1 + \frac{75}{100}\frac{\cos(70y)}{12}\right) + \frac{\cos(100y)^2}{24} + 4xy$$

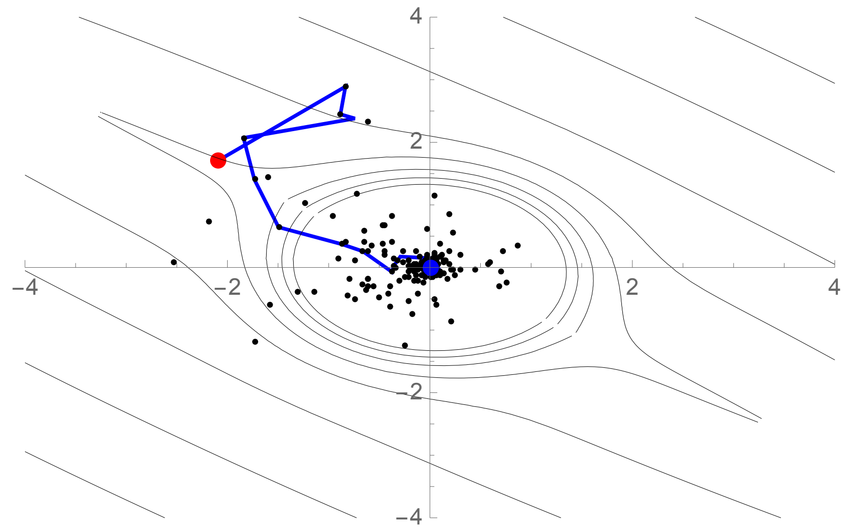


Log plot, missing values are numerical inconsistencies.

Canoe function.

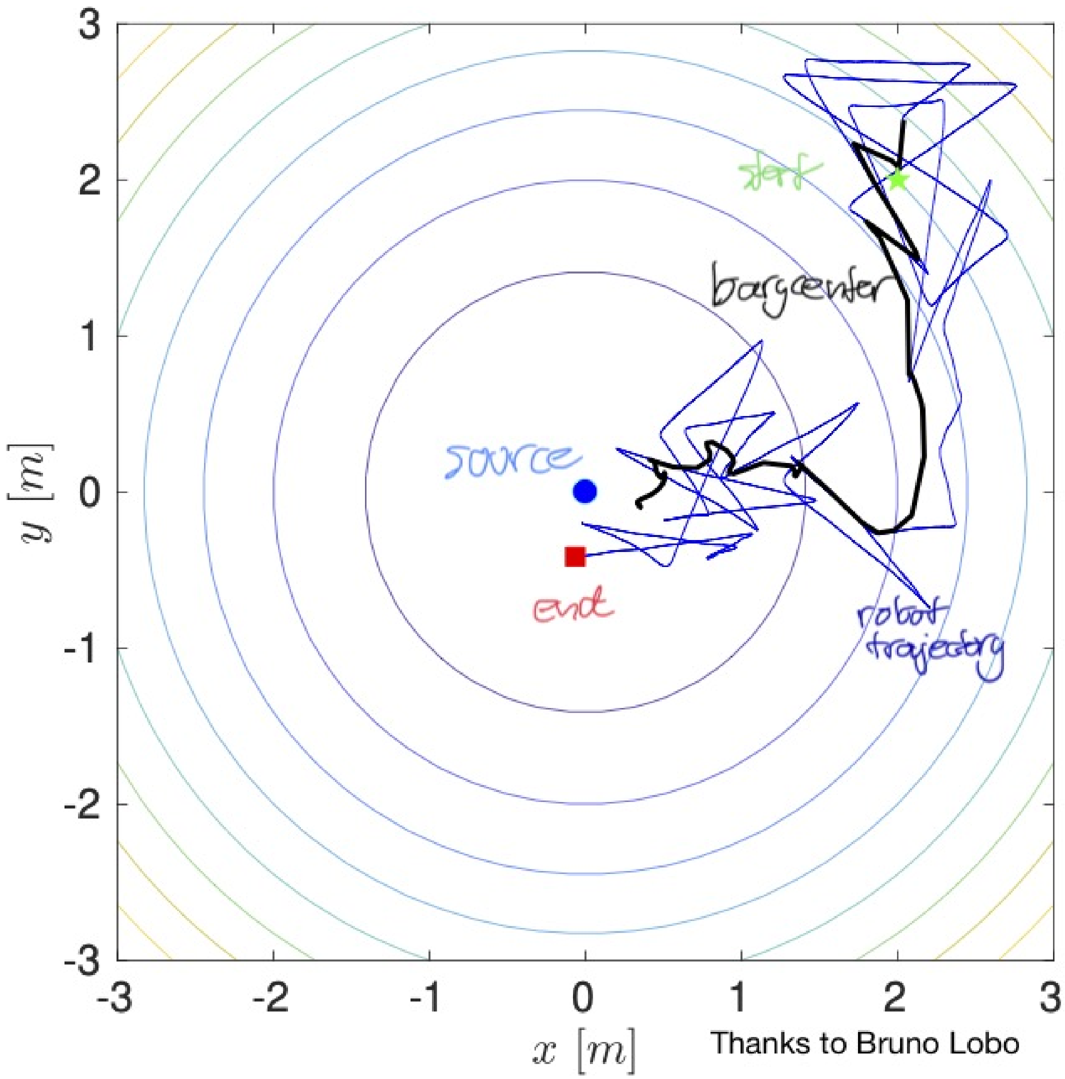
$(1 - e^{||x||^2}) \max(||x - c||^2, ||x - d||^2)$, with $c = -d = [30, 40]^T$





(Apologies for the obsolete data visualization.)

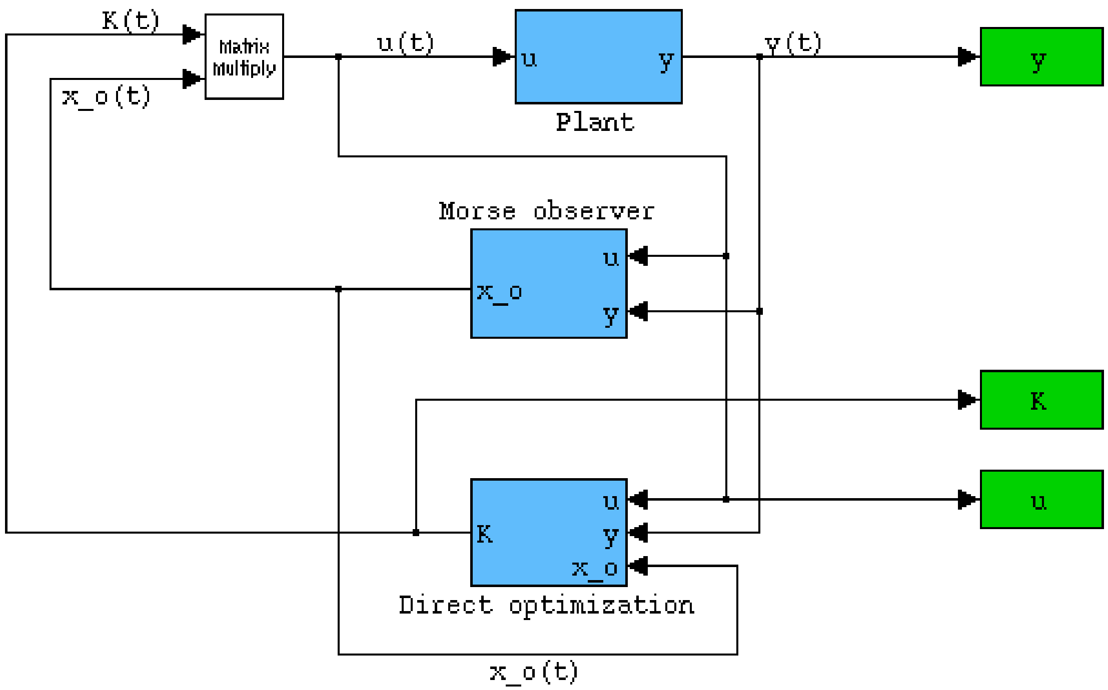
Application: Robot searching signal source.



In the spirit of extremum seeking control.

Work by Bruno Lobo, USP.

Adaptive control example.



Open–loop unstable, non-minimum phase plant:

$$G(s) = \frac{0.01(3 - 8s)}{s^2 - 0,01}.$$

Work by Rodrigo Romano, Instituto Mauá de Tecnologia.

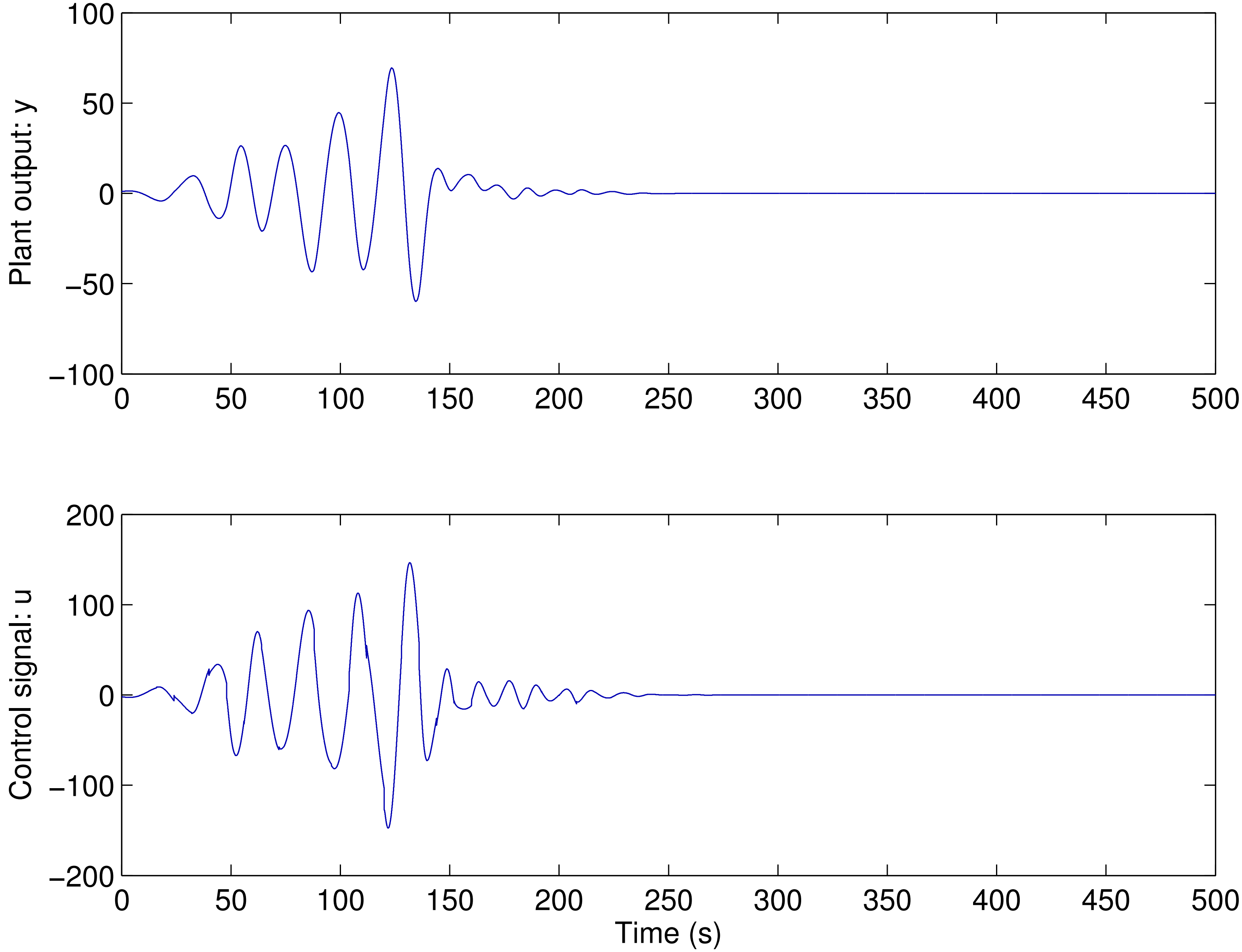
Simulation example.

Morse observer with 2 pole pairs at $s = -1.414(1 \pm j)$.

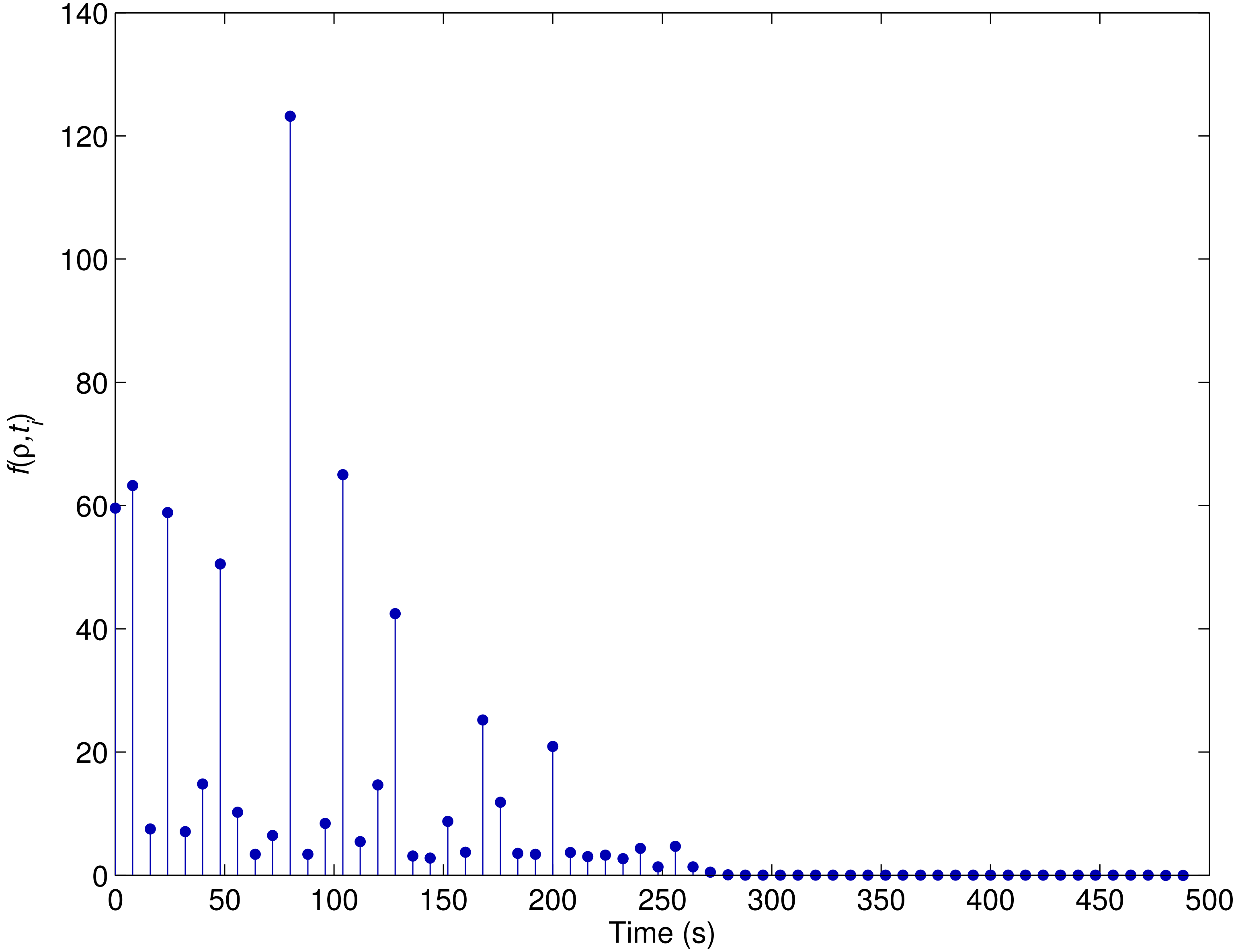
Direct optimization settings:

- $T = 8s$ (duration of intervals).
- $|z|^2 = y^T Q y + u^T R u$, with $Q = 10$ and $R = 0.1$.
- $K(t_0) = -\,[2 \;\; 1.5 \;\; 0 \;\; 0]$ (closed–loop unstable with initial gain).

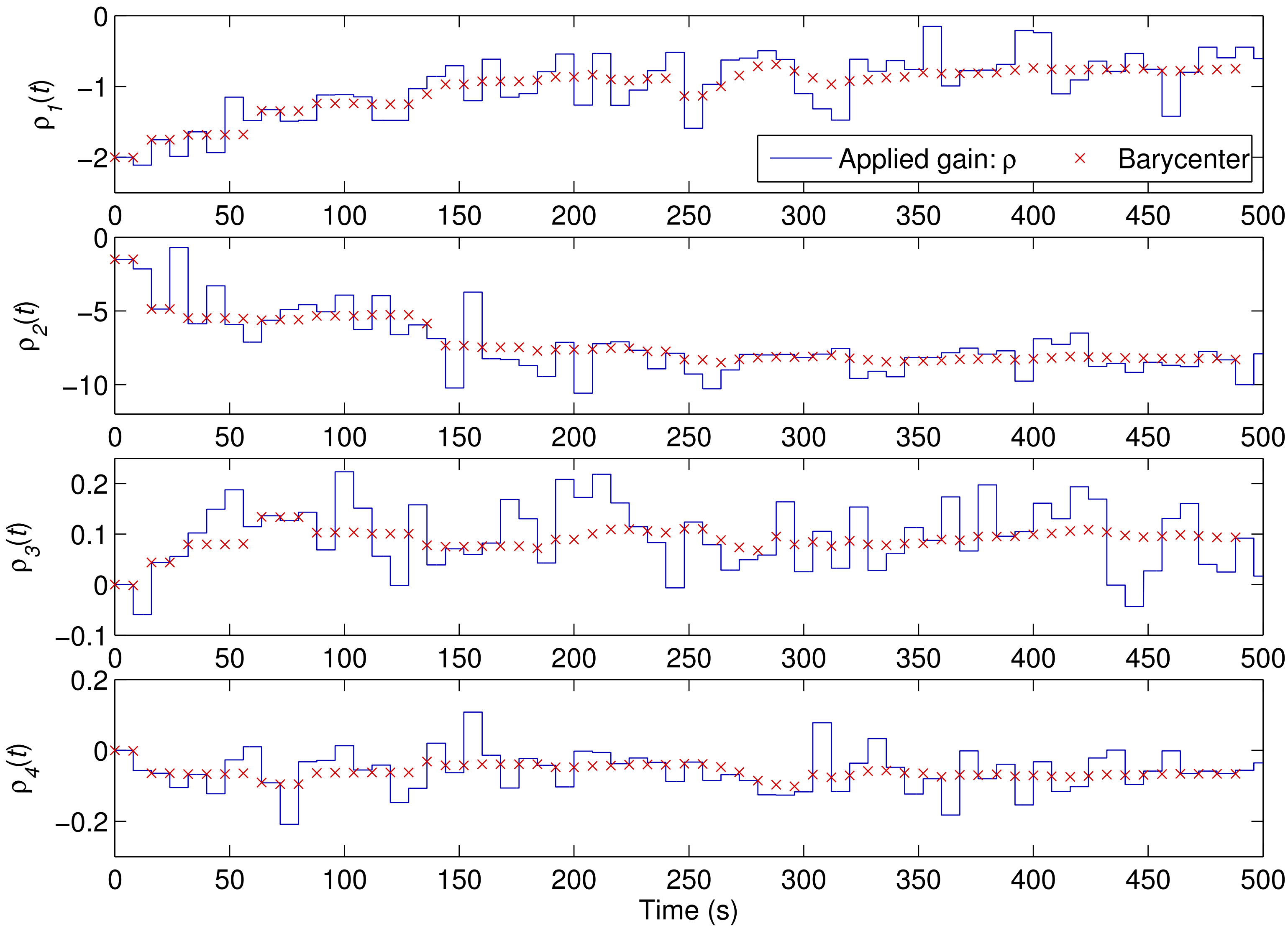
Plant input output behaviour.



Barycenter objective function value.



Feedback controller parameters.



Morse observer.

Control plant with output $y \in \mathbb{R}^{n_y}$, input $u \in \mathbb{R}^{n_u}$ using observer–based direct adaptive controller

$$\dot{x} = A_{\mathcal{O}} x + B_{\mathcal{O}} u + D_{\mathcal{O}} y \tag{11}$$

$$u = K(t) x. \tag{12}$$

$A_{\mathcal{O}}$, $B_{\mathcal{O}}$, $D_{\mathcal{O}}$ are fixed, K is a tunable feedback parameter. State and output estimators

$$\widehat{x} = E_{\mathcal{O}}(\theta) x \tag{13}$$

$$\widehat{y} = C_{\mathcal{O}}(\theta) x + G_{\mathcal{O}}(\theta) y. \tag{14}$$

Quadratic error equation.

Compute

$$|z|^2 = y^T Q y + u^T R u. \tag{15}$$

Stabilizability of $(A, B) \Rightarrow \exists P > 0$ such that

$$A^T P + PA - PBR^{-1}B^T P + C_{\mathcal{O}}^T (I - G)^{-1} Q (I - G)^{-1} C_{\mathcal{O}} = 0$$

for positive–definite matrices Q and R.

$$\boxed{\int_{t_i-T}^{t_i} |z|^2\,dt = \int_{t_i-T}^{t_i} x^T (K - K_*)^T R (K - K_*) x dt}$$

$$+x(t_i-T)^T Px(t_i-T) - x(t_i)^T Px(t_i)$$

Direct optimization.

Feedback K linearly parametrized by ρ.
Task: tune ρ to minimize

$$\int_{t_i-T}^{t_i} |z|^2 \, dt$$

Solution exists: $K(\rho) = K_*$.
Gradient methods not available.

During each interval $[t_i - T, t_i)$, apply constant feedback $K(\rho_i)$, obtain the control cost $f(\rho_i)$.

At each t_i use barycenter formula to compute $\widehat{\rho}(t_i)$.

Pick $\rho(t_i)$ as a random variable with mean $\widehat{\rho}(t_i)$.

Estimate parameters of hybrid linear model.

$$\begin{aligned}\dot{x}(t) &= A_\sigma^M x(t) + B_\sigma^M u(t)\\ y(t) &= C_\sigma^M x(t) + D_\sigma^M u(t),\end{aligned}$$

using measurements $\{u(t_k), y(t_k)\}$, $k \in \{0, 1, \ldots, N\}$ of input and output signals.

Switching signal $\sigma(t) \in \{1, 2, \ldots, S\}$ is piecewise constant between switching times $\{\tau_1, \ldots, \tau_S\}$, which themselves are unknown a priori and need to be determined.

Also work with Rodrigo Romano, Instituto Mauá de Tecnologia.

Consider single model switch, use regressor form.

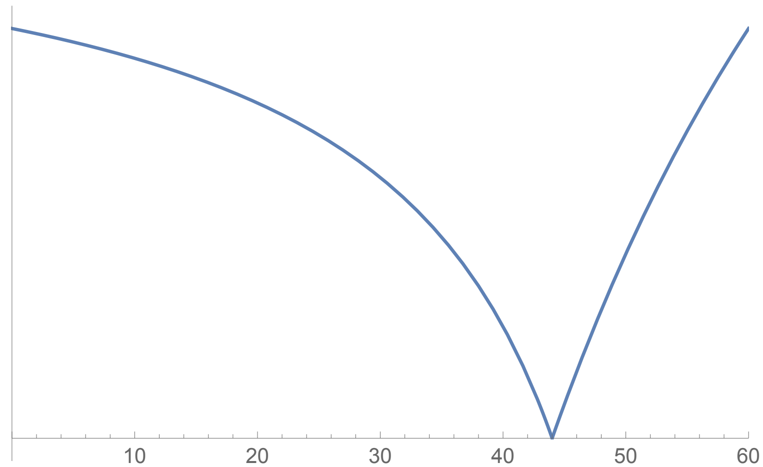
$$\widehat{y}(t) = \theta_{\sigma}^{T}\phi(t),$$

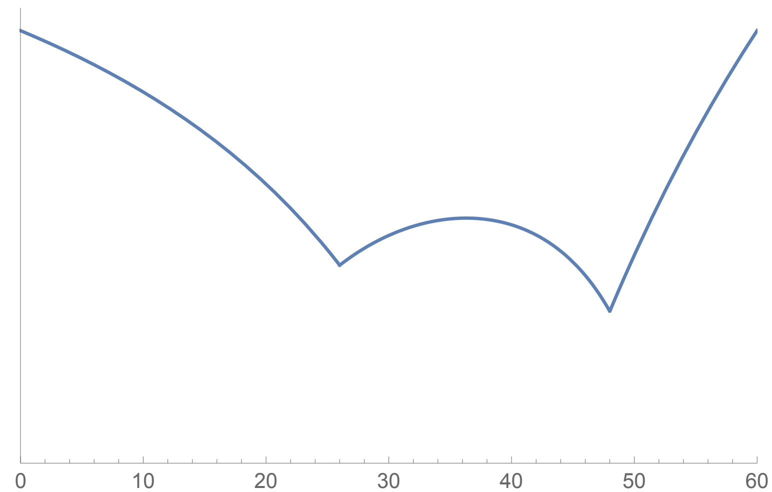
Search for candidate switching time τ that minimizes integral square error $J(\tau) = J_1(\tau) + J_2(\tau)$ with

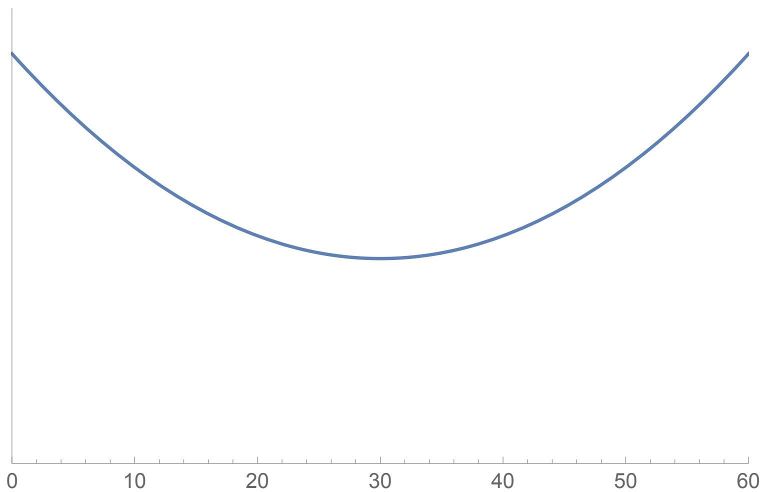
$$J_1(\tau) = \int_{t_0}^{\tau} |\widehat{\theta}_1^T(\tau)\phi(s) - y(s)|^2 \, \mathrm{d}s$$
$$J_2(\tau) = \int_{\tau}^{t_N} |\widehat{\theta}_2^T(\tau)\phi(s) - y(s)|^2 \, \mathrm{d}s$$

Given τ, obtain expressions for $\widehat{\theta}_1^T(\tau)$ and $\widehat{\theta}_2^T(\tau)$ (least–squares).

Rough sketch of cost functional.







Once–switching, twice–switching, and randomly drifting actual plants.

medialunas porteñas

Not convex, even for single switching.
Derivative–free optimization or gradient–descent possible, 2nd order methods not effective.

The barycenter method is

- Easy to implement. Complex version at least seems new.
- Amenable to analysis: equivalence between batch and recursive versions.
- Naturally parallelizable (combined center of mass).
- Flexible: can incorporate other techniques.
- Applicable to nonconvex, noisy, not differentiable functions.
- Not too shabby.

varying ν for multifidelity?

What do you think?

Thanks!

Hoping for an in person visit at an appropriate time.