Two applications of discrete variational calculus

David Martín de Diego (ICMAT)

Contact Mechanics and its neighbours

A mini symposium in occasion of the graduation of Federico Zadra

October 2, 2023

First application:

Parallel approach to the solutions of Discrete Euler-Lagrange equations

Lagrangian system

- Lagrangian system: (Q, L) [see e.g. Abraham & Marsden (1978)]
  - Q, configuration manifold;
  - $L : TQ \to \mathbb{R}$, (continuous) Lagrangian.

Lagrangian system

- Lagrangian system: (Q, L) [see e.g. Abraham & Marsden (1978)]
  - Q, configuration manifold;
  - $L : TQ \to \mathbb{R}$, (continuous) Lagrangian.
- Hamilton's principle: Actual trajectories $c \in \mathcal{C}^2(q_0, q_1, [t_0, t_1])$ of the system are critical points of the action $\mathcal{S} : C^2(q_0, q_1, [t_0, t_1]) \to \mathbb{R}$,

$$\mathcal{S}[c] = \int_{t_0}^{t_1} L(c^{(1)}(t))\,\mathrm{d}t = \int_{t_0}^{t_1} L(q^i(t), \dot{q}^i(t))\,\mathrm{d}t$$

Lagrangian system

- Lagrangian system: (Q, L) [see e.g. Abraham & Marsden (1978)]
  - Q, configuration manifold;
  - $L : TQ \to \mathbb{R}$, (continuous) Lagrangian.
- Hamilton's principle: Actual trajectories $c \in \mathcal{C}^2(q_0, q_1, [t_0, t_1])$ of the system are critical points of the action $\mathcal{S} : C^2(q_0, q_1, [t_0, t_1]) \to \mathbb{R}$,
$$\mathcal{S}[c] = \int_{t_0}^{t_1} L(c^{(1)}(t))\,\mathrm{d}t = \int_{t_0}^{t_1} L(q^i(t), \dot{q}^i(t))\,\mathrm{d}t$$
- Critical iff Euler-Lagrange eqs. (EL) satisfied:
$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))\right) - \frac{\partial L}{\partial q^i}(q(t), \dot{q}(t)) = 0, \quad i = 1, \ldots, n = \dim Q$$

Lagrangian system

- Lagrangian system: (Q, L) [see e.g. Abraham & Marsden (1978)]
  - Q, configuration manifold;
  - $L : TQ \to \mathbb{R}$, (continuous) Lagrangian.
- Hamilton's principle: Actual trajectories $c \in \mathcal{C}^2(q_0, q_1, [t_0, t_1])$ of the system are critical points of the action $\mathcal{S} : C^2(q_0, q_1, [t_0, t_1]) \to \mathbb{R}$,
$$\mathcal{S}[c] = \int_{t_0}^{t_1} L(c^{(1)}(t))\,\mathrm{d}t = \int_{t_0}^{t_1} L(q^i(t), \dot{q}^i(t))\,\mathrm{d}t$$
- Critical iff Euler-Lagrange eqs. (EL) satisfied:
$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))\right) - \frac{\partial L}{\partial q^i}(q(t), \dot{q}(t)) = 0, \quad i = 1, \dots, n = \dim Q$$
- Legendre transformation: $\mathbb{F}L : TQ \to T^*Q$, locally $(q^i, \dot{q}^i) \mapsto \left(q^i, p_i := \frac{\partial L}{\partial \dot{q}^i}(q, \dot{q})\right)$. If $\mathbb{F}L$ local (global) isomorphism, L regular (hyperregular)$\longrightarrow$ Hamiltonian description

Discrete mechanical systems

Discrete Lagrangian problem

- Discrete Lagrangian system: (Q, L_d) [Marsden & West (2001)]
  - Q, configuration manifold;
  - $L_d : Q \times Q \to \mathbb{R}$, discrete Lagrangian.

Discrete mechanical systems

Discrete Lagrangian problem

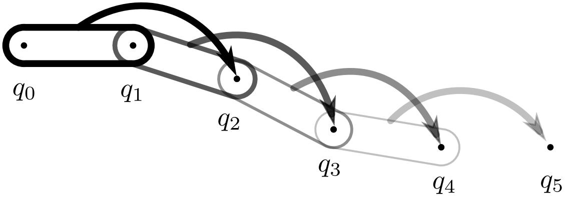
- Discrete Lagrangian system: (Q, L_d) [Marsden & West (2001)]
  - Q, configuration manifold;
  - $L_d : Q \times Q \to \mathbb{R}$, discrete Lagrangian.
- Variational problem: Find critical $c_d := \{q_k \in Q\}_{k=0}^{N}$ of $\mathcal{S}_d[c_d] := \sum_{k=0}^{N-1} L_d(q_k, q_{k+1})$ with fixed q_0, q_N.

Discrete mechanical systems

Discrete Lagrangian problem

- Discrete Lagrangian system: (Q, L_d) [Marsden & West (2001)]
  - Q, configuration manifold;
  - $L_d : Q \times Q \to \mathbb{R}$, discrete Lagrangian.
- Variational problem: Find critical $c_d := \{q_k \in Q\}_{k=0}^{N}$ of $\mathcal{S}_d[c_d] := \sum_{k=0}^{N-1} L_d(q_k, q_{k+1})$ with fixed q_0, q_N.
- Critical iff discrete Euler-Lagrange eqs (DEL) satisfied:

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0, \quad k = 1, ..., N-1$$



Discrete mechanical systems

Discrete Lagrangian problem

- Discrete Lagrangian system: (Q, L_d) [Marsden & West (2001)]
  - Q, configuration manifold;
  - $L_d : Q \times Q \to \mathbb{R}$, discrete Lagrangian.
- Variational problem: Find critical $c_d := \{q_k \in Q\}_{k=0}^{N}$ of $\mathcal{S}_d[c_d] := \sum_{k=0}^{N-1} L_d(q_k, q_{k+1})$ with fixed q_0, q_N.
- Critical iff discrete Euler-Lagrange eqs (DEL) satisfied:

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0, \quad k = 1, ..., N-1$$

Regularity

A discrete Lagrangian, $L_d : Q \times Q \to \mathbb{R}$, is said to be **regular** if its associated block matrix $\mathcal{W}_d = D_{12} L_d = \left(\frac{\partial^2 L_d}{\partial q_0 \partial q_1} \right)$ is regular.

Discrete fibre derivatives

$$
\begin{array}{rrcl}
\mathbb{F}^{-}L_d : & Q \times Q & \to & T^*Q \\
 & (q_0, q_1) & \mapsto & (q_0, p_0 := -D_1 L_d(q_0, q_1)) \\
\mathbb{F}^{+}L_d : & Q \times Q & \to & T^*Q \\
 & (q_0, q_1) & \mapsto & (q_1, p_1 := D_2 L_d(q_0, q_1))
\end{array}
$$

If these are local isomorphisms, then L_d is said to be regular.

Discrete fibre derivatives

$$
\begin{array}{rrcl}
\mathbb{F}^- L_d : & Q \times Q & \to & T^*Q \\
& (q_0, q_1) & \mapsto & (q_0, p_0 := -D_1 L_d(q_0, q_1)) \\
\mathbb{F}^+ L_d : & Q \times Q & \to & T^*Q \\
& (q_0, q_1) & \mapsto & (q_1, p_1 := D_2 L_d(q_0, q_1))
\end{array}
$$

If these are local isomorphisms, then L_d is said to be regular.

Equivalent to $D_{12} L_d(q_0, q_1)$ regular.

Discrete fibre derivatives

$$
\begin{array}{rrcl}
\mathbb{F}^- L_d : & Q \times Q & \to & T^*Q \\
& (q_0, q_1) & \mapsto & (q_0, p_0 := -D_1 L_d(q_0, q_1)) \\
\mathbb{F}^+ L_d : & Q \times Q & \to & T^*Q \\
& (q_0, q_1) & \mapsto & (q_1, p_1 := D_2 L_d(q_0, q_1))
\end{array}
$$

If these are local isomorphisms, then L_d is said to be regular.

Equivalent to $D_{12} L_d(q_0, q_1)$ regular.

Discrete flows

Discrete Lagrangian flow: $F_{L_d} : Q \times Q \to Q \times Q$, induced by DEL.

Discrete Hamiltonian flow: $\widetilde{F}_{L_d} : T^*Q \to T^*Q$,
$\widetilde{F}_{L_d} = \mathbb{F}^\pm L_d \circ F_{L_d} \circ (\mathbb{F}^\pm L_d)^{-1}$.

Symplecticity $(\widetilde{F}_{L_d})^* \omega_Q = \omega_Q$.

Exact discrete Lagrangian

Relation with continuous Lagrangian problems

$L_d(q_k, q_{k+1}) \approx \int_0^h L(q(t), \dot{q}(t))\mathrm{d}t$, fixed $h \in \mathbb{R}$, q solution of continuous Euler-Lagrange equations s.t. $q(0) = q_k$, $q(h) = q_{k+1}$.

$$L_d^e(q_0, q_1) = \int_0^h L(q(t), \dot{q}(t))\, dt,$$

where $q(t)$ is a trajectory of the continuous system joining q_0 to q_1 for time h. If L is regular, then L_d^e regular.
If $q(t)$ is a solution of the continuous system, then the evolution of the discrete system for L_d^e yields the sequence $q(0)$, $q(h)$, $q(2h)$, $q(3h)$, ...

Exact Discrete Lagrangian problem

Marsden-West 2001

Let $L_d\colon Q \times Q \to \mathbb{R}$ be a discrete Lagrangian. We say that L_d is a discretization of order r if there exist an open subset $U_1 \subset TQ$ with compact closure and constants $C_1 > 0$, $h_1 > 0$ so that

$$|L_d(q(0), q(h)) - L_d^e(q(0), q(h))| \leq C_1 h^{r+1}$$

for all solutions $q(t)$ of the second-order Euler–Lagrange equations with initial conditions $(q_0, \dot{q}_0) \in U_1$ and for all $h \leq h_1$.

Exact Discrete Lagrangian problem

Marsden-West 2001

Let $L_d \colon Q \times Q \to \mathbb{R}$ be a discrete Lagrangian. We say that L_d is a discretization of order r if there exist an open subset $U_1 \subset TQ$ with compact closure and constants $C_1 > 0$, $h_1 > 0$ so that

$$|L_d(q(0), q(h)) - L_d^e(q(0), q(h))| \leq C_1 h^{r+1}$$

for all solutions $q(t)$ of the second-order Euler–Lagrange equations with initial conditions $(q_0, \dot{q}_0) \in U_1$ and for all $h \leq h_1$.

Theorem [Patrick & Cuell, 2009]

If L is a regular Lagrangian and L_d is a discrete Lagrangian for L of order r, then

$$\widetilde{F}_{L_d} = \widetilde{F}_{L_d^e} + \mathcal{O}(h^{r+1})$$

γ-th order Lagrangian problems

(Q, L) with $L : T^{(\gamma)}Q \to \mathbb{R}$ Lagrangian function.

Point $q^{[\gamma]} \in T^{(\gamma)}Q$, local coords. $(q, \dot{q}, ..., q^{(\gamma)})$.

Higher order Euler-Lagrange equations (ODEs of order 2γ):

$$\sum_{\alpha=0}^{\gamma}(-1)^{\alpha}\frac{d^{\alpha}}{dt^{\alpha}}\left(\frac{\partial L}{\partial q^{(\alpha)i}}\right) = 0, \quad i = 1, ..., \dim Q$$

Fibre derivative:

$$\begin{array}{rccc} \mathbb{F}L\ : & T^{(2\gamma-1)}Q & \to & T^{*}T^{(\gamma-1)}Q \\ & (q^{i}, ..., q^{(2\gamma-1)\,i}) & \mapsto & (q^{i}, ..., q^{(\gamma-1)\,i}, p_{i,0}, ..., p_{i,\gamma-1}) \end{array}$$

with $p_{i,\alpha} = \sum_{\beta=0}^{\gamma-\alpha-1}(-1)^{\beta}\frac{d^{\beta}}{dt^{\beta}}\left(\frac{\partial L}{\partial q^{(\beta+\alpha+1)}{}_{i}}\right)$ Jacobi-Ostrogradski momenta.

Example: Second order Lagrangian problem

- Lagrangian system: $(T^{(2)}Q, L)$

Example: Second order Lagrangian problem

- Lagrangian system: $(T^{(2)}Q, L)$
- Euler-Lagrange equations:

$$\frac{\partial L}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^i}\right) + \frac{\mathrm{d}^2}{\mathrm{d}^2 t}\left(\frac{\partial L}{\partial \ddot{q}^i}\right) = 0, \quad i = 1, ..., n$$

Higher order discrete mechanical systems

For the γ-th order case, a discrete Lagrangian is given as a function

$$L_d \colon T^{(\gamma-1)}Q \times T^{(\gamma-1)}Q \to \mathbb{R}$$

Higher order discrete mechanical systems

For the γ-th order case, a discrete Lagrangian is given as a function

$$L_d\colon T^{(\gamma-1)}Q \times T^{(\gamma-1)}Q \to \mathbb{R}$$

The is a sum

$$\sum_{k=0}^{N-1} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]})$$

Higher order discrete mechanical systems

For the γ-th order case, a discrete Lagrangian is given as a function

$$L_d \colon T^{(\gamma-1)}Q \times T^{(\gamma-1)}Q \to \mathbb{R}$$

The is a sum

$$\sum_{k=0}^{N-1} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]})$$

The condition that a sequence $\{q_k^{[\gamma-1]}\}_{k=0}^{N}$ of points in $T^{(\gamma-1)}Q$ be critical for the discrete action, with fixed endpoints $q_0^{[\gamma-1]}$ and $q_N^{[\gamma-1]}$, is equivalent to the equations

$$D_2 L_d(q_{k-1}^{[\gamma-1]}, q_k^{[\gamma-1]}) + D_1 L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}) = 0$$

(DEL equations, order γ).

Second order discrete mechanical systems

Second order discrete Lagrangian problem

[Colombo, Ferraro, MdD (2016)]

TQ, configuration manifold;

- $\circ$ $L_d : TQ \times TQ \to \mathbb{R}$, discrete Lagrangian.

Variational problem: Find critical $c_d := \{(q_k, v_k) \in TQ\}_{k=0}^{N}$ of $\mathcal{S}_d[c_d] := \sum_{k=0}^{N-1} L_d(q_k, v_k, q_{k+1}, v_{k+1})$ with fixed (q_0, v_0), (q_N, v_N).

- Critical iff discrete Euler-Lagrange eqs (DEL) satisfied:

$$D_3 L_d(q_{k-1}, v_{k-1}, q_k, v_k) + D_1 L_d(q_k, v_k, q_{k+1}, v_{k+1}) = 0, k = 1, ..., N-1$$

$$D_4 L_d(q_{k-1}, v_{k-1}, q_k, v_k) + D_2 L_d(q_k, v_k, q_{k+1}, v_{k+1}) = 0.$$

Regularity

A discrete Lagrangian, $L_d : T^{(\gamma-1)}Q \times T^{(\gamma-1)}Q \to \mathbb{R}$, is said to be **regular** if its associated block matrix $\mathcal{W}_d = \left(\dfrac{\partial^2 L_d}{\partial q_0^{[\gamma-1]} \partial q_1^{[\gamma-1]}} \right)$ or

$$\mathcal{W}_d = \begin{pmatrix} \frac{\partial^2 L_d}{\partial q_0 \partial q_1} & \frac{\partial^2 L_d}{\partial q_0 \partial \dot{q}_1} & \cdots & \frac{\partial^2 L_d}{\partial q_0 \partial q_1^{(\gamma-1)}} \\ \frac{\partial^2 L_d}{\partial \dot{q}_0 \partial q_1} & \frac{\partial^2 L_d}{\partial \dot{q}_0 \partial \dot{q}_1} & \cdots & \frac{\partial^2 L_d}{\partial \dot{q}_0 \partial q_1^{(\gamma-1)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L_d}{\partial q_0^{(\gamma-1)} \partial q_1} & \frac{\partial^2 L_d}{\partial q_0^{(\gamma-1)} \partial \dot{q}_1} & \cdots & \frac{\partial^2 L_d}{\partial q_0^{(\gamma-1)} \partial q_1^{(\gamma-1)}} \end{pmatrix}$$

is regular.

Exact discrete Lagrangian

Starting from a continuous Lagrangian L, we define the exact discrete Lagrangian as

$$L_d^e(q_0^{[\gamma-1]}, q_1^{[\gamma-1]}) = \int_0^h L(q^{[\gamma]}(t))\, dt$$

where $q\colon [0, h] \to Q$ is the unique $C^{2\gamma}$ solution curve of the Euler-Lagrange equations satisfying the boundary conditions $q^{[\gamma-1]}(0) = q_0^{[\gamma-1]}$ and $q^{[\gamma-1]}(h) = q_1^{[\gamma-1]}$.

Exact discrete Lagrangian

Starting from a continuous Lagrangian L, we define the exact discrete Lagrangian as

$$L_d^e(q_0^{[\gamma-1]}, q_1^{[\gamma-1]}) = \int_0^h L(q^{[\gamma]}(t))\, dt$$

where $q\colon [0,h] \to Q$ is the unique $C^{2\gamma}$ solution curve of the Euler-Lagrange equations satisfying the boundary conditions $q^{[\gamma-1]}(0) = q_0^{[\gamma-1]}$ and $q^{[\gamma-1]}(h) = q_1^{[\gamma-1]}$. This exact discrete Lagrangian is well-defined for h small enough and in a neighborhood U_h of the diagonal of $T^{(\gamma-1)}Q \times T^{(\gamma-1)}Q$. We also know that U_h degenerates into the diagonal for $h = 0$.

Boundary value problems

The boundary value problem posed by these equations, i.e.

$$\text{given fixed } q_0^{[\gamma-1]}, q_N^{[\gamma-1]} \text{ find } c_d^* \text{ s.t. } c_d^*(0) = q_0^{[\gamma-1]}, c_d^*(N) = q_N^{[\gamma-1]}$$

$$\begin{array}{rcl}
D_2L_d(q_0^{[\gamma-1]}, q_1^{[\gamma-1]}) + D_1L_d(q_1^{[\gamma-1]}, q_2^{[\gamma-1]}) & = & 0, \\
D_2L_d(q_1^{[\gamma-1]}, q_2^{[\gamma-1]}) + D_1L_d(q_2^{[\gamma-1]}, q_3^{[\gamma-1]}) & = & 0, \\
\dots & = & \dots \\
D_2L_d(q_{N-2}^{[\gamma-1]}, q_{N-1}^{[\gamma-1]}) + D_1L_d(q_{N-1}^{[\gamma-1]}, q_N^{[\gamma-1]}) & = & 0
\end{array}$$

This problem can be solved using different strategies, e.g.

This problem can be solved using different strategies, e.g.

- Initial value methods
  - Simple shooting (Can be highly unstable)
  - Multiple shooting (Better alternative. Admits parallelization)

This problem can be solved using different strategies, e.g.

- Initial value methods
  - Simple shooting (Can be highly unstable)
  - Multiple shooting (Better alternative. Admits parallelization)
- Global or finite difference methods (Can require a lot of memory and computational power)

This problem can be solved using different strategies, e.g.

- Initial value methods
  - Simple shooting (Can be highly unstable)
  - Multiple shooting (Better alternative. Admits parallelization)
- Global or finite difference methods (Can require a lot of memory and computational power)

Our approach is a **parallelized relaxation approach** based on the latter methods.

Parallel algorithm

Problem divided in independent sub-problems. Outputs are combined to form desired output.

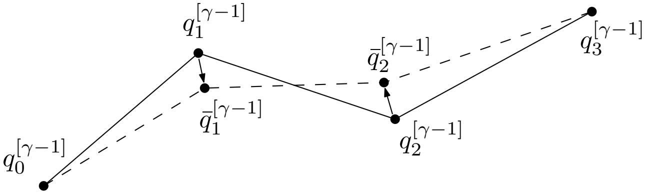


Figure: An iteration of the parallel method, for $N = 3$.

Find a sequence $\{q_k\}_{k=0}^N$ that is a solution of DEL.

Parallelized Discrete Euler-Lagrange equations

- Choose initial guess c_d^0, such that $c_d^0(0) = q_0$, $c_d^0(N) = q_N$;
- Find c_d^ℓ satisfying $c_d^\ell(0) = q_0$, $c_d^\ell(N) = q_N$ and

$$D_2 L_d(q_{k-1}^{\ell-1}, q_k^\ell) + D_1 L_d(q_k^\ell, q_{k+1}^{\ell-1}) = 0 \quad k = 1, ..., N-1, \ell = 1, 2, ...$$

Convergence

Root-finding problem: $f = (f_1, ..., f_n) : \mathbb{R}^n \to \mathbb{R}^n$, find $x^* \in \mathbb{R}^n$ s.t. $f(x^*) = 0$. We may parallelize the problem as follows:

Algorithm. Nonlinear Jacobi method

- Choose initial guess $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$;
- Find x^ℓ satisfying

$$f(x_1^1, ..., x_{k-1}^{\ell-1}, x_k^\ell, x_{k+1}^{\ell-1}, ..., x_n^{\ell-1}) = 0, \quad k = 1, ..., N-1,\ \ell = 1, 2, ...$$

$$\begin{aligned} f_1({\color{red}x_1^1}, x_2^0, \ldots, x_n^0) &= 0, \\ f_2(x_1^0, {\color{red}x_2^1}, ..., x_n^0) &= 0, \\ \ldots &= 0 \\ f_n(x_1^0, x_2^0, \ldots, {\color{red}x_n^1}) &= 0 \end{aligned}$$

Convergence

Theorem. Convergence of the Jacobi process [see Vrahatis (2003)]

Let $F : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open neighborhood $\mathcal{S}_0 \subset \mathcal{D}$ of a point $x^* \in \mathcal{D}$ for which $f(x^*) := \nabla F(x^*) = 0$, and suppose that the Hessian $H(x^*)$ of F is positive definite, block-tridiagonal with regular blocks on the diagonal. Then there exists an open ball $\mathcal{S} \subset \mathcal{S}_0$ centered at x^* such that any sequence $\{x^j\}_{j=0}^{\infty}$, $x^0 \in \mathcal{S}$, generated by the nonlinear Jacobi process converges to x^*.

Given N, $(t_a, q_a) = (t_0, q_0)$, $(t_b, q_b) = (t_N, q_N)$, find a sequence $\mathring{c}_d^* := \{(t_k, q_k)\}_{k=1}^{N-1}$ with $t_{k+1} - t_k = h$ that is a solution of DEL.

Parallelized Discrete Euler-Lagrange equations

Choose initial guess $\mathring{c}_d^0$;

Find $\mathring{c}_d^\ell$ satisfying

$$D_2 L_d(q_{k-1}^{\ell-1}, q_k^\ell) + D_1 L_d(q_k^\ell, q_{k+1}^{\ell-1}) = 0 \quad k = 1, ..., N-1, \ell = 1, 2, ...$$

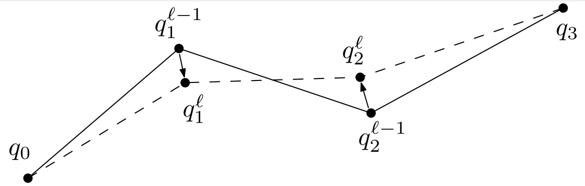
Given N, $(t_a, q_a) = (t_0, q_0)$, $(t_b, q_b) = (t_N, q_N)$, find a sequence $\mathring{c}_d^* := \{(t_k, q_k)\}_{k=1}^{N-1}$ with $t_{k+1} - t_k = h$ that is a solution of DEL.

Parallelized Discrete Euler-Lagrange equations

Choose initial guess $\mathring{c}_d^0$;

Find $\mathring{c}_d^\ell$ satisfying

$$D_2 L_d(q_{k-1}^{\ell-1}, q_k^\ell) + D_1 L_d(q_k^\ell, q_{k+1}^{\ell-1}) = 0 \quad k = 1, ..., N-1, \ell = 1, 2, ...$$



Given N, $(t_a, q_a) = (t_0, q_0)$, $(t_b, q_b) = (t_N, q_N)$, find a sequence $\mathring{c}_d^* := \{(t_k, q_k)\}_{k=1}^{N-1}$ with $t_{k+1} - t_k = h$ that is a solution of DEL.

Parallelized Discrete Euler-Lagrange equations

Choose initial guess $\mathring{c}_d^0$;

Find $\mathring{c}_d^\ell$ satisfying

$$D_2 L_d(q_{k-1}^{\ell-1}, q_k^\ell) + D_1 L_d(q_k^\ell, q_{k+1}^{\ell-1}) = 0 \quad k = 1, \ldots, N-1, \ell = 1, 2, \ldots$$

Convergence can be accelerated by bundling several nodes together at the expense of increased load per thread, i.e. instead of solving $(q_{k-1}^{\ell-1}, q_k^\ell, q_{k+1}^{\ell-1})$, solve $(q_{k-a}^{\ell-1}, q_{k-a+1}^\ell, \ldots, q_k^\ell, \ldots, q_{k+b-1}^\ell, q_{k+b}^{\ell-1})$ with $a + b \geq 2$.

Ingredients

$\mathring{c}_d^*$ acts as our x^*;

$\mathcal{J}_{L_d}[c_d^*]$ acts as our $F(x^*)$;

$\nabla F(x^*)$ are the DEL.

We need to show that the Hessian $H(x^*)$ of F is positive definite.

Boundary value problems in mechanics

The discrete equations for L_d are $\nabla F = 0$ where F is the discrete action $\sum_{k=0}^{N-1} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]})$ as a function of $x = \mathbf{q}^{[\gamma-1]} = (q_1^{[\gamma-1]}, \ldots, q_{N-1}^{[\gamma-1]})$, and $q_0^{[\gamma-1]}$ and $q_N^{[\gamma-1]}$ are fixed. Since L_d is C^2, then the Hessian of F is symmetric and has the block tridiagonal form

$$H(\mathbf{q}^{[\gamma-1]}) = \begin{bmatrix} D_1 & C_1 & & & \\ C_1^T & D_2 & C_2 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{N-3}^T & D_{N-2} & C_{N-2} \\ & & & C_{N-2}^T & D_{N-1} \end{bmatrix}$$

$$D_k = D_{22}L_d(q_{k-1}^{[\gamma-1]}, q_k^{[\gamma-1]}) + D_{11}L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}), \quad k = 1, \ldots, N-1,$$

$$C_k = D_{12}L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}), \quad k = 1, \ldots, N-2.$$

Convergence

Jacobi convergence

Let $\mathbf{q}^{[\gamma-1]*} = (q_1^{[\gamma-1]*}, \ldots, q_{N-1}^{[\gamma-1]*})$ be a solution of the DEL equations for fixed $q_0^{[\gamma-1]}$ and $q_N^{[\gamma-1]}$. If the Hessian of the discrete action, $H(\mathbf{q}^{[\gamma-1]*})$, is positive definite, then the block Jacobi method converges locally to $\mathbf{q}^{[\gamma-1]*}$.

Convergence

$$H(\mathbf{q}^{[\gamma-1]}) = \begin{pmatrix} B_0 + A_1 & C_1 & & & \\ C_1^\top & B_1 + A_2 & C_2 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{N-3}^\top & B_{N-3} + A_{N-2} & C_{N-2} \\ & & & C_{N-2}^\top & B_{N-2} + A_{N-1} \end{pmatrix}$$

where

$$\begin{aligned} A_k &= D_{11} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}), \quad & k = 1, \ldots, N-1, \\ B_k &= D_{22} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}), \quad & k = 0, \ldots, N-2, \\ C_k &= D_{12} L_d(q_k^{[\gamma-1]}, q_{k+1}^{[\gamma-1]}), \quad & k = 1, \ldots, N-2, \end{aligned}$$

and $q_0^{[\gamma-1]}$, $q_N^{[\gamma-1]}$ are fixed.

Theorem

Denote by $\mathcal{D}_i = \mathcal{B}_{i-1} + \mathcal{A}_i$, $1 \leq i \leq N-1$.
If the matrices defined iteratively by $\Lambda_1 = \mathcal{D}_1 = \mathcal{B}_0 + \mathcal{A}_1$ and

$$\Lambda_i = \mathcal{D}_i - C_{i-1}^T \Lambda_{i-1}^{-1} C_{i-1}, \quad 2 \leq i \leq N-1$$

are all positive definite then the Hessian matrix $H((q^{[\gamma-1]}))$ is positive definite.

However, we want more!

Relevant questions

If L_d is a discretization of L, what can we say about H_{L_d}?

Is H_{L_d} positive-definite if $D_{22}L$ positive-definite?

Is H_{L_d} even regular if L is regular?

However, we want more!

Relevant questions

If L_d is a discretization of L, what can we say about H_{L_d}?

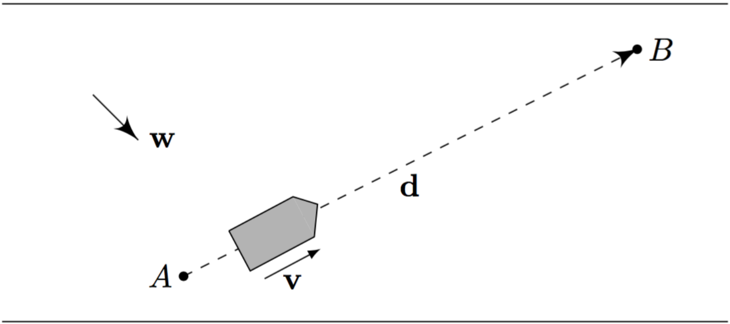
- Is H_{L_d} positive-definite if $D_{22}L$ positive-definite?
- Is H_{L_d} even regular if L is regular?

These are not so immediate to answer. Non-local!

Jacobi equations and conjugate points (continuous and discrete). Work in progress

Zermelo's navigation problem

From (Zermelo, 1931), (Bao et al., 2004), (Javaloyes, Sánchez, 2017), (Kopacz, 2019) and more...



Zermelo's navigation problem

Statement

Time-optimal control problem: Find the minimum time (ship) trajectories γ on a Riemannian manifold (Q, g) under the influence of a drift vector field (wind) $W \in \mathfrak{X}(Q)$. Assume $\|\dot{\gamma}(s) - W(\gamma(s))\|_g = 1$ and $\alpha(q) := 1 - \|W(q)\|_g > 0$ for all $q \in Q$.

These minimum time trajectories are geodesics for a Randers metric:

$$F(q, v_q) = \sqrt{a(v_q, v_q)} + \langle b(q), v_q \rangle \, ,$$

where

$$a(v_q, v_q) := \frac{1}{\alpha(q)} g(v_q, v_q) + \langle b(q), v_q \rangle^2,$$

$$\langle b(q), v_q \rangle := -\frac{1}{\alpha(q)} g(W(q), v_q).$$

Zermelo's navigation problem II

The time it takes the ship to move along a curve $\gamma : [s_0, s_N] \to Q$ is

$$t[\gamma] = \int_{s_0}^{s_N} F(\gamma(s), \dot{\gamma}(s))\, \mathrm{d}s.$$

Zermelo's navigation problem II

The time it takes the ship to move along a curve $\gamma : [s_0, s_N] \to Q$ is

$$t[\gamma] = \int_{s_0}^{s_N} F(\gamma(s), \dot{\gamma}(s))\, \mathrm{d}s.$$

The action functional

$$\mathcal{S}[\gamma] = \int_{s_0}^{s_N} L(\gamma(s), \dot{\gamma}(s))\, \mathrm{d}s := \int_{s_0}^{s_N} F(\gamma(s), \dot{\gamma}(s))^2\, \mathrm{d}s\,,$$

defines a regular Lagrangian whose extremals will coincide with time-extremal curves.

Example

- $Q = \mathbb{R}^2$, Euclidean metric.
- $W = 1.7 \cdot (-R_{2,2} - R_{4,4} - R_{2,5} + R_{5,1})$, where

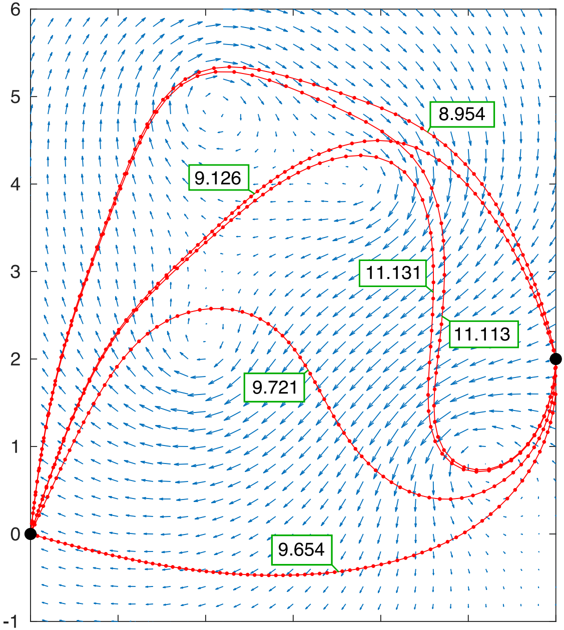
$$R_{a,b}(x,y) = \frac{1}{3((x-a)^2 + (y-b)^2) + 1} \begin{bmatrix} -y+b \\ x-a \end{bmatrix}$$

- Discrete Lagrangian:

$$L_d(q_0, q_1) = \frac{h}{2} \left[F\left(q_0, \frac{q_1 - q_0}{h}\right)^2 + F\left(q_1, \frac{q_1 - q_0}{h}\right)^2 \right].$$

- Boundary conditions: $q_0 = (0,0)$, $q_{80} = (6,2)$.

Zermelo's navigation problem



Fuel-optimal navigation problem

A related but inequivalent variant of the problem:

Statement

Let $T > 0$ be a fixed time. Find trajectories from $(x(0), y(0))$ to $(x(T), y(T))$ minimizing the cost functional

$$\mathcal{S}[u] = \int_0^T \frac{1}{2}(u_1^2 + u_2^2)\, \mathrm{d}t\,,$$

subject to

$$\begin{aligned}\dot{x} &= u_1 + W_1(x, y)\,,\\ \dot{y} &= u_2 + W_2(x, y)\,.\end{aligned}$$

This problem is equivalent to solving the Euler-Lagrange equations for the Lagrangian

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\left[(\dot{x} - W_1(x, y))^2 + (\dot{y} - W_2(x, y))^2\right]\,.$$

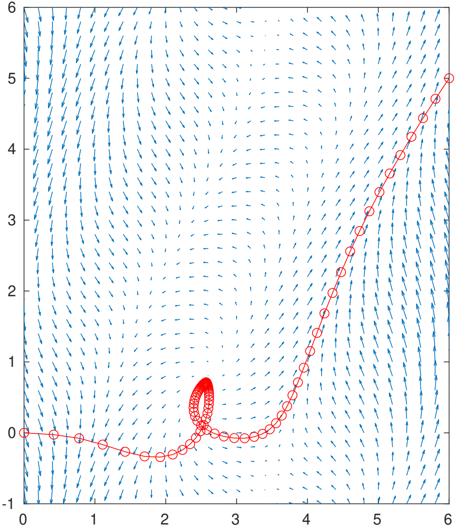
Example

- $Q = \mathbb{R}^2$, Euclidean metric.
- $W = (\cos(2x - y - 6), 2/3 \sin(y) + x - 3)$.
- Discrete Lagrangian:

$$L_d(q_0, q_1) = \frac{h}{2}\left[L\left(q_0, \frac{q_1 - q_0}{h}\right) + L\left(q_1, \frac{q_1 - q_0}{h}\right)\right].$$

- $N = 200$.
- Total navigation time $T = 30$.
- Boundary conditions: $q(0) = (0, 0)$, $q(T) = (6, 5)$.

Example



Smooth fuel-optimal navigation problem

Statement

Let $T > 0$ be a fixed time. Find trajectories passing through given points $\{q(t_i)\}_{i=1}^m$, $0 = t_0 < \ldots < t_i < \ldots < t_m = T$, with $\dot{q}(0)$ and $\dot{q}(T)$ fixed, minimizing the cost functional

$$\mathcal{S}[u] = \int_0^T \frac{1}{2}\left[u_1^2 + u_2^2 + c(v_1^2 + v_2^2)\right]\, \mathrm{d}t\,,$$

with $c > 0$ subject to

$$\begin{array}{rclcrcl} \dot{x} & = & u_1 + W_1(x,y), & \quad & \dot{y} & = & u_2 + W_2(x,y), \\ \dot{u}_1 & = & v_1, & & \dot{u}_2 & = & v_2\,. \end{array}$$

Equivalent to solving the Euler-Lagrange equations of

$$\begin{aligned} L(x,y,\dot{x},\dot{y},\ddot{x},\ddot{y}) = \frac{1}{2}\Big[&(\dot{x} - W_1(x,y))^2 + (\dot{y} - W_2(x,y))^2 \\ &+ c\,(\ddot{x} - D_1 W_1(x(t),y(t))\dot{x} - D_2 W_1(x(t),y(t))\dot{y})^2 \\ &+ c\,(\ddot{y} - D_1 W_2(x(t),y(t))\dot{x} - D_2 W_2(x(t),y(t))\dot{y})^2\Big] \end{aligned}$$

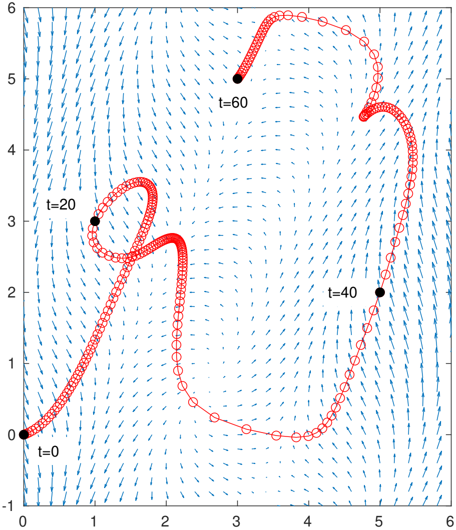
Example

- $Q = \mathbb{R}^2$, Euclidean metric.
- W same as in the former example.
- Discrete Lagrangian:

$$
\begin{aligned}
L_d(q_0, v_0, q_1, v_1) = \frac{h}{2} \Bigg[& L\left(q_0, v_0, \frac{2}{h^2}(3(q_1 - q_0) - h(v_1 + 2v_0)\right) \\
& + L\left(q_1, v_1, -\frac{2}{h^2}(3(q_1 - q_0) - h(2v_1 + v_0)\right) \Bigg]
\end{aligned}
$$

- $c = 50$, $N = 240$.
- Total navigation time $T = 60$.
- Boundary conditions: $(q(0), v(0)) = (0, 0, 0, 0)$, $(q(T), v(T)) = (3, 5, 0, 0)$.
- Interpolation conditions: $q(T/3) = (1, 3)$, $q(2T/3) = (5, 2)$.

Example



- Conclusions:
  - Discrete variational methods combined with a parallel iterative approach are well-suited for boundary value problems.
  - These give us alternatives to multiple shooting and are suited for GPU implementation.
  - Tested in three examples related with navigation problems.
  - Can be readily extended to the Lie group setting.
- Outlook:
  - Handling of equality constraints.
  - Handling of inequality constraints via penalty potentials and coupling with iteration progress.
  - Application to new examples (astrodynamics, time-dependent flows...)

Real application

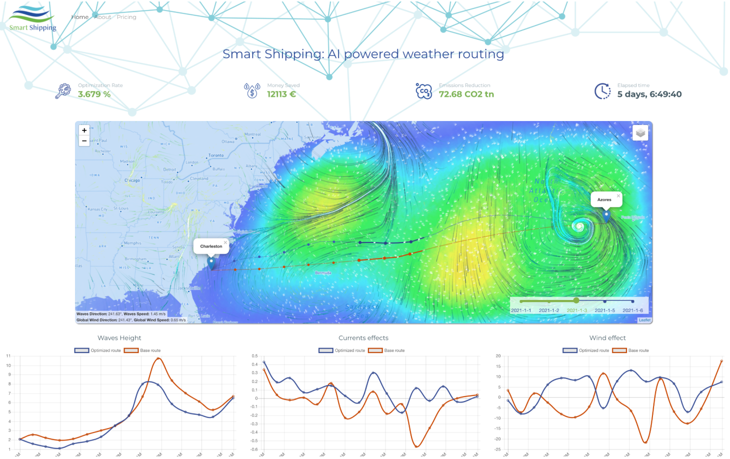


Figure: **Prototype of web application for the Smart Shipping weather routing project. Red : original route. Blue : optimized route, consuming 3.7% less fuel and reducing 72 tons of GHG emissions.**

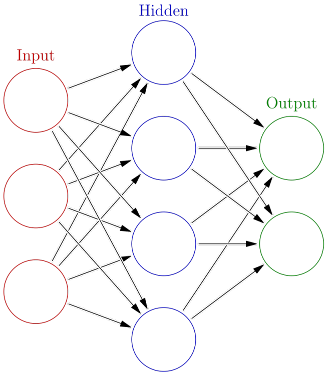
References

- M. N. Vrahatis, G. D. Magoulas and V. P. Plagianakos. From linear to nonlinear iterative methods. *Appl. Numer. Math.* (2003);
- E. Zermelo. Über das Navigationsproblem bei ruhender oder veränderlicher Windverteilung.*Z. Angew. Math. Mech.* (1931);
- D. Bao, C. Robles and Z. Shen. Zermelo navigation on Riemannian manifolds. *J. Differential Geom.* (2004);
- P. Kopacz On generalization of Zermelo navigation problem on Riemannian manifolds. *Int. J. Geom. Methods Mod. Phys.* (2019);
- M. A. Javaloyes, M. Sánchez. Wind Riemannian spaceforms and Randers-Kropina metrics of constant flag curvature. *Eur. J. Math.* (2017);
- L. Colombo, S. Ferraro and DMdD. Geometric integrators for higher-order variational systems and their application to optimal control. *J. Nonlinear Sci.* (2016);
- S. Ferraro, DMdD, Rodrigo Sato. Parallel iterative methods for variational integration applied to navigation problems. IFAC PapersOnLine 54-19, 321–326. (2021);
- J.E. Marsden & M. West Discrete Mechanics and Variational Integrators. *Acta Num.* (2001);

Second application: accelerated optimization

Motivation

$\mathcal{X}$ input space
$\mathcal{Y}$ output space
Θ Parameter space
A map $\Psi : \mathcal{X} \times \Theta \longrightarrow \mathcal{Y}$ is called a **neural network**



Motivation

Modern statistical data analysis involves very large data sets and very large parameter spaces, so that **computational efficiency** is very importance in practical applications.

In large-scale data analysis, in many cases algorithms need to be linear, or nearly linear, in relevant problem parameters.

Motivation

For a given finite set of pairs $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ (**training data**), try to determine parameters $\theta^* \in \Theta$ such that

$$\Psi(x_i, \theta^*) \approx y_i$$

At the end, a **neural network** is a function (perhaps) consisting of thousands or millions of parameters, that represents a mathematical solution to a real problem.

Motivation

The function $\Psi : \mathcal{X} \times \Theta \longrightarrow \mathcal{Y}$ typically consists of a composition of S-**layers**: $\{\psi^0, \psi^1, \psi^{S-1}\}$

$$\Psi = \psi^{S-1} \circ \ldots \circ \psi^1 \circ \psi^0$$

where $\psi^s : \mathcal{X}^k \times \Theta^k \longrightarrow \mathcal{X}^{k+1}$, where $\mathcal{X}^0 = \mathcal{X}$ and $\mathcal{X}^S = \mathcal{Y}$.

Loss function:

$$\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$

for instance $\mathcal{L}(y, y^*) = \frac{1}{2}\|y - y^*\|^2$

Given the training data $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ $1 \leq i \leq N$ (**training data**) define the **Total Loss**

$$\min_{\theta \in \Theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\Psi(x_i, \theta), y_i) + R(\theta) \right\}$$

R is a regularizer penalizing unwanted parameter solutions.

Loss function:

$$\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$

for instance $\mathcal{L}(y, y^*) = \frac{1}{2}\|y - y^*\|^2$

Given the training data $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ $1 \leq i \leq N$ (**training data**) define the **Total Loss**

$$\min_{\theta \in \Theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\Psi(x_i, \theta), y_i) + R(\theta) \right\}$$

R is a regularizer penalizing unwanted parameter solutions.

$$\underset{x \in D}{\arg\min} f(x)$$

Optimizers are used to solve optimization problems by minimizing the function!

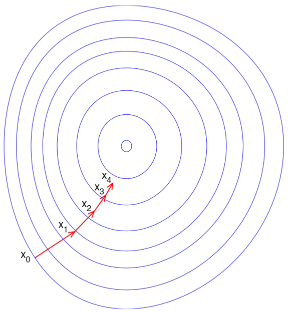
Is there an optimal way to optimize?

Gradient descend

$$\arg\min_{x \in D} f(x)$$

$$x_{k+1} = x_k - \epsilon \nabla f(x_k)$$

The idea is to take repeated steps in the opposite direction of the gradient of f at the current point, because this is the direction of steepest descent.



However, the iterates may converge slowly. Converge to the optimum at the rate $\mathcal{O}(1/k)$, where k is the number of iterations.

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

Simple Test Function

Rosenbrock function, 1960

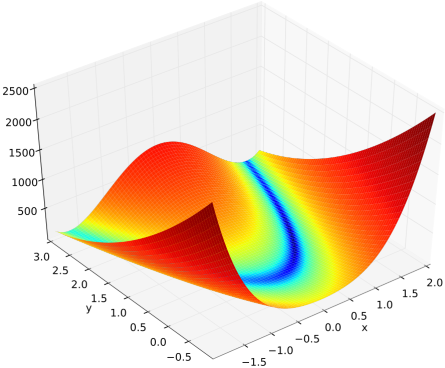
$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$

Steep well, flat valley

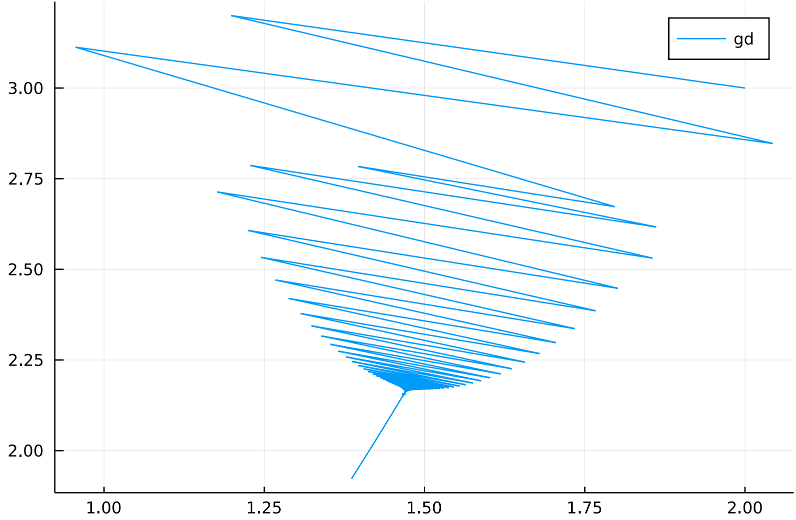
Banana shaped

Global minimum at (a, a^2)

$$f(a, a^2) = 0$$



Gradient descent



[] []

In

> *A.Nemirovsky and D.Yudin,* **Problem complexity and method efficiency in optimization Problem**, *ser. Interscience Series in Discrete Mathematics. John Wiley, 1983.*

proved that no first-order method can converge at a rate faster than $\mathcal{O}(1/k^2)$ on convex optimization problems with Lipschitz-continuous gradient.

Accelerated optimization

In 1983,

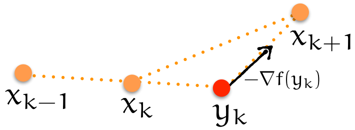
> *Y. Nesterov,* **A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$,** *Soviet Mathematics Doklady, vol. 27, pp. 372–376, 1983.*

introduced a new method, **Nesterov Accelerated Gradient** (NAG), that further improved the convergence rate.

Nesterov Accelerated Gradient

$$x^* = \arg\min_{x \in D} f(x)$$

$$x_{k+1} = y_k - \epsilon \nabla f(y_k)$$
$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$$



Convergence rate $\mathcal{O}(1/k^2)$
Oscillatory but faster

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

(SU, Boyd, Candes '16)

Optimization meets Geometric Mechanics

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

W. Su, S. Boyd, and E. J. Candès, **A differential equation for modeling nesterov'saccelerated gradient method: Theory and insights,** Journal of Machine Learning Research, 17 (2016), pp. 1–43.

Optimization meets Geometric Mechanics

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

W. Su, S. Boyd, and E. J. Candès, **A differential equation for modeling nesterov'saccelerated gradient method: Theory and insights,** Journal of Machine Learning Research, 17 (2016), pp. 1–43.

Euler-Lagrange equations

$$L(x, \dot{x}, t) = t^3 \left(\frac{1}{2}\dot{x} - f(x) \right)$$

A. Wibisono, A. C. Wilson, and M. I. Jordan, **A variational perspective on accelerated methods in optimization**, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. E7351–E7358

"Such a variational perspective also has the advantage of being generative—we can derive algorithms that achieve fast rates rather than requiring an analysis to establish a fast rate for a specific algorithm that is derived in an adhoc manner"...

Michael I. Jordan DYNAMICAL, SYMPLECTIC AND STOCHASTIC PERSPECTIVES ON GRADIENT-BASED OPTIMIZATION, Proceedings of the Internation Congress of Mathematiciens – 2018 Rio de Janeiro, Vol. 1 (523–550)

“Such a variational perspective also has the advantage of being generative—we can derive algorithms that achieve fast rates rather than requiring an analysis to establish a fast rate for a specific algorithm that is derived in an adhoc manner”...

Michael I. Jordan DYNAMICAL, SYMPLECTIC AND STOCHASTIC PERSPECTIVES ON GRADIENT-BASED OPTIMIZATION, Proceedings of the Internation Congress of Mathematiciens – 2018 Rio de Janeiro, Vol. 1 (523–550)

...“we will find that symplectic integrators, which are widely used for integrating dynamics obtained from variational or Hamiltonian perspectives, are relevant in the optimization setting”

Bregman Lagrangians

Define a **Bregman divergence** :

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle d\Phi(y), x - y \rangle \ .$$

Φ is a convex distance-generating function
As a typical example, if $\Phi(x) = \frac{1}{2}\|x\|^2$ then

$$\mathcal{B}_\Phi(x, y) = \frac{1}{2}\|x - y\|^2 \ .$$

From a Bergman divergence we can construct the Bregman kinetic energy $K\colon \mathbb{R} \times T\mathbb{R}^n \to \mathbb{R}$ by

$$K(x, v, t) = \mathcal{B}_\Phi(x + e^{-\alpha(t)}v, x)$$

and the potential energy

$$U(x, t) = e^{\beta(t)} f(x)$$

to then define the **Bregman Lagrangian**:

$$L(x, v, t) = e^{\alpha(t)+\gamma(t)} \left(K(x, v, t) - U(x, t)\right)$$

where the time-dependent functions $\alpha(t), \beta(t), \gamma(t)$ are chosen to produce different algorithms.

Time-dependent mechanics

Let Q be a manifold and TQ its tangent bundle. As usual, coordinates (x^i) on Q induce coordinates $(x^i, \dot{x}^i)$ on TQ. Therefore we have natural coordinates $(x^i, \dot{x}^i, t)$ on $TQ \times \mathbb{R}$ which is the appropriate velocity phase space for time-dependent systems.

Time-dependent mechanics

Let Q be a manifold and TQ its tangent bundle. As usual, coordinates (x^i) on Q induce coordinates $(x^i, \dot{x}^i)$ on TQ. Therefore we have natural coordinates $(x^i, \dot{x}^i, t)$ on $TQ \times \mathbb{R}$ which is the appropriate velocity phase space for time-dependent systems. Let $a, b \in \mathbb{R}$ with $a < b$, given two points $x_a, x_b \in Q$, we consider the set of curves:

$$C^2([a,b], x_a, x_b) = \{\sigma \colon [a,b] \to Q \mid \sigma \in C^2 \text{ with } \sigma(a) = x_a,\ \sigma(b) = x_b\}$$

Given a **time-dependent Lagrangian** function $L \colon TQ \times \mathbb{R} \to \mathbb{R}$ define the action $\mathcal{J}_L \colon C^2([a,b], x_a, x_b) \to \mathbb{R}$ by

$$\mathcal{J}_L(\sigma) = \int_a^b L(\sigma'(t), t)\, dt \tag{1}$$

where $\sigma' \colon [a,b] \to TQ$ is defined by $\sigma'(t) = \frac{d\sigma}{dt}(t) \in T_{\sigma(t)}Q$.

Time-dependent mechanics

Euler-Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) - \frac{\partial L}{\partial x^i} = 0, \quad 1 \leq i \leq n = \dim Q \tag{2}$$

$$\frac{dE_L}{dt} = \frac{\partial L}{\partial t}\,.$$

Time-dependent mechanics

$$\mathcal{F}L\colon TQ \times \mathbb{R} \to T^*Q \times \mathbb{R}$$

$$\mathcal{F}L(x^i, \dot{x}^i, t) = (x^i, \frac{\partial L}{\partial \dot{x}^i}, t)$$

Take induced coordinates (x^i, p_i, t) on $T^*Q \times \mathbb{R}$.
We assume that the Legendre transformation is a diffeomorphism (that is, the Lagrangian is hyperregular) and define the Hamiltonian function $H\colon T^*Q \times \mathbb{R} \to \mathbb{R}$ by:

$$H = E_L \circ (\mathcal{F}L)^{-1} \; .$$

Define the projections $\mathrm{pr}_1 \colon T^*Q \times \mathbb{R} \to T^*Q$ and $\mathrm{pr}_2 \colon T^*Q \times \mathbb{R} \to \mathbb{R}$ we induce the **cosymplectic structure** (Ω_H, η) on $T^*Q \times \mathbb{R}$ where

$$\eta = \mathrm{pr}_2^* dt \ , \qquad \Omega_H = -d(\mathrm{pr}_1^*\theta_Q - H\eta) = \Omega_Q + dH \wedge dt \ .$$

Here θ_Q denotes the Liouville 1-form on T^*Q given in induced coordinates by $\theta_Q = p_i \, dx^i$. Denote by $\Omega_Q = -d\mathrm{pr}_1^*\theta_Q$ the pullback of the canonical symplectic 2-form $\omega_Q = -d\theta_Q$ on T^*Q. In coordinates, $\Omega_Q = dx^i \wedge dp_i$ (observe that now Ω_Q is presymplectic since $\ker \Omega_Q = \mathrm{span}\ \{\partial/\partial t\}$). Therefore in induced coordinates (x^i, p_i, t):

$$\Omega_H = dx^i \wedge dp_i + dH \wedge dt \ , \qquad \eta = dt$$

Time-dependent mechanics

We define the **evolution vector field** $E_H \in \mathfrak{X}(T^*Q \times \mathbb{R})$ by

$$i_{E_H}\Omega_H = 0 \,, \qquad i_{E_H}\eta = 1$$

In local coordinates the evolution vector field is:

$$E_H = \frac{\partial}{\partial t} + \frac{\partial H}{\partial p_i}\frac{\partial}{\partial x^i} - \frac{\partial H}{\partial x^i}\frac{\partial}{\partial p_i} \,.$$

The integral curves of E_H are given by:

$$\dot{t} = 1 \,, \qquad \dot{x}^i = \frac{\partial H}{\partial p_i} \,, \qquad \dot{p}_i = -\frac{\partial H}{\partial x^i} \,.$$

The integral curves of E_H are precisely the curves of the form $t \to \mathcal{F}L(\sigma'(t), t)$ where $\sigma \colon I \to Q$ is a solution of the Euler-Lagrange equations for $L \colon TQ \times \mathbb{R} \to \mathbb{R}$.

Time-dependent mechanics. An example

If we consider the Nesterov Lagrangian function $L\colon T\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$

$$L(x, \dot{x}, t) = t^3 \left(\frac{1}{2} \|\dot{x}\|^2 - f(x) \right) . \tag{3}$$

The Legendre transformation is $\mathcal{F}L(x, \dot{x}, t) = (x, p = t^3 \dot{x}, t)$ and the Hamiltonian function

$$H(x, p, t) = \frac{1}{2t^3} \|p\|^2 + t^3 f(x)$$

In this case the Hamilton equations are:

$$\dot{t} = 1 , \qquad \dot{x} = \frac{p}{t^3} , \qquad \dot{p} = -t^3 \nabla f(x) . \qquad \diamond$$

Time-dependent mechanics

$$\mathcal{L}_{E_H}(\Omega_Q + dH \wedge dt) = 0 \qquad \mathcal{L}_{E_H}\eta = 0$$

The flow of the evolution vector field $\Psi_s \colon \mathcal{U} \subset T^*Q \times \mathbb{R} \to T^*Q \times \mathbb{R}$

$$\Psi_s(\alpha_q, t) = (\Psi_{t,s}(\alpha_q), t+s), \qquad \alpha_q \in T^*_q Q,$$

Therefore from the flow of E_H we induce a map

$$\Psi_{t,s} \colon \mathcal{U}_t \subseteq T^*Q \to T^*Q$$

where $\mathcal{U}_t = \{\alpha_q \in T^*Q \mid (\alpha_q, t) \in \mathcal{U}\}$.

$$\Psi_s^*(\Omega_Q + dH \wedge dt) = \Omega_Q + dH \wedge dt, \qquad \Psi_s^*(\eta) = \eta \ .$$

Time-dependent mechanics

$\Psi_s : \mathcal{U} \subset T^*Q \times \mathbb{R} \to T^*Q \times \mathbb{R}$

Theorem: We have that $\Psi_{t,s} : \mathcal{U}_t \subseteq T^*Q \to T^*Q$ is a **symplectomorphism**, that is, $\Psi^*_{t,s}\omega_Q = \omega_Q$.

Discrete variational methods for time-dependent Lagrangian systems

Consider $Q \times Q$ as a discrete version of TQ and, instead of curves on Q, the solutions are replaced by sequences of points on Q.

$$\mathcal{C}_d(Q) = \left\{ x_d \colon \{k\}_{k=0}^{N} \to Q \right\}$$

for the set of possible sequences, which can be identified with the manifold $Q \times \overset{(N+1)}{\cdots} \times Q$.

Discrete variational methods for time-dependent Lagrangian systems

Consider $Q \times Q$ as a discrete version of TQ and, instead of curves on Q, the solutions are replaced by sequences of points on Q.

$$\mathcal{C}_d(Q) = \left\{ x_d \colon \{k\}_{k=0}^{N} \to Q \right\}$$

for the set of possible sequences, which can be identified with the manifold $Q \times \overset{(N+1)}{\cdots} \times Q$.

A **discrete time-dependent Lagrangian** is a family of maps

$$L_d^k \colon Q \times Q \to \mathbb{R}, \qquad k = 0, \dots, N-1 \, .$$

Discrete variational methods for time-dependent Lagrangian systems

Define the **discrete action map**, on the space of sequences $\mathcal{C}_d(Q)$ by

$$S_d(x_d) = \sum_{k=0}^{N-1} L_d^k(x_k, x_{k+1}), \quad x_d \in \mathcal{C}_d(Q)\,.$$

If we consider variations of x_d with fixed end points x_0 and x_N and extremize S_d over $x_1, \ldots, x_{N-1}$, we obtain the **discrete Euler-Lagrange equations** (DEL for short)

$$\partial_{x_{k+1}} S_d(x_d) = D_1 L_d^{k+1}(x_{k+1}, x_{k+2}) + D_2 L_d^k(x_k, x_{k+1}) = 0 \;\; k = 0, \ldots, N-2\,.$$

Define a discrete flow $F_{k,k+1}\colon\, TQ \to TQ$

Discrete variational methods for time-dependent Lagrangian systems

If L_d^k is regular for all k, that is, the matrix

$$D_{12}L_d^k = \left(\frac{\partial^2 L_d^k}{\partial x_k \partial x_{k+1}}\right)$$

is non-singular, the two discrete Legendre transformations associated to L_d^k

$$\mathbb{F}^+L_d^k, \mathbb{F}^-L_d^k \colon Q \times Q \to T^*Q\,, k = 1, \ldots, N$$

by

$$\begin{aligned}\mathbb{F}^+L_d^k &\colon (x_k, x_{k+1}) \longmapsto (x_{k+1}, D_2L_d^k(x_k, x_{k+1}))\,,\\ \mathbb{F}^-L_d^k &\colon (x_k, x_{k+1}) \longmapsto (x_k, -D_1L_d^k(x_k, x_{k+1}))\,.\end{aligned}$$

are local diffeomorphisms.

Discrete variational methods for time-dependent Lagrangian systems

We can also define the evolution of the discrete system on the Hamiltonian side, $\tilde{F}_{k,k+1}\colon T^*Q \to T^*Q$, by any of the formulas

$$
\begin{aligned}
\tilde{F}_{k,k+1} &= \mathbb{F}^+ L_d^k \circ (\mathbb{F}^- L_d^k)^{-1} = \mathbb{F}^+ L_d^k \circ F_{k-1,k} \circ (\mathbb{F}^+ L_d^{k-1})^{-1} \\
&= \mathbb{F}^- L_d^{k+1} \circ F_{k,k+1} \circ (\mathbb{F}^- L_d^k)^{-1} \,,
\end{aligned}
$$

because of the commutativity of the following diagram:

$$
\begin{tikzcd}
(x_{k-1}, x_k) \arrow[rr, "F_{k-1,k}"] \arrow[dr, "\mathbb{F}^+ L_d^{k-1}"'] & & (x_k, x_{k+1}) \arrow[rr, "F_{k,k+1}"] \arrow[dl, "\mathbb{F}^- L_d^k"'] \arrow[dr, "\mathbb{F}^+ L_d^k"] & & (x_{k+1}, x_{k+2}) \arrow[dl, "\mathbb{F}^- L_d^{k+1}"'] \\
& (x_k, p_k) \arrow[rr, "\tilde{F}_{k,k+1}"'] & & (x_{k+1}, p_{k+1}) &
\end{tikzcd}
$$

Discrete variational methods for time-dependent Lagrangian systems

The discrete Hamiltonian map $\tilde{F}_{k,k+1} \colon (T^*Q, \omega_Q) \to (T^*Q, \omega_Q)$ is a **symplectic transformation**, that is

$$(\tilde{F}_{k,k+1})^* \omega_Q = \omega_Q \; .$$

Discrete variational methods for time-dependent Lagrangian systems.Examples

The most simple discretization of the discrete Lagrangian is given by approximating the action using the initial point:

$$L_{d,h}^{k,\mathrm{ini}}(x_k, x_{k+1}) = hL(x_k, \frac{x_{k+1} - x_k}{h}, kh)$$

In the case of the Nesterov Lagrangian $L(x, \dot{x}, t) = t^3 \left(\frac{1}{2}\|\dot{x}\|^2 - f(x)\right)$ the corresponding first-order discrete Euler-Lagrange equations for the Nesterov Lagrangian is:

$$x_{k+2} - x_k = \frac{k^3}{(1+k)^3}(x_{k+1} - x_k) - h^2 \nabla f(x_{k+1})$$

Discrete variational methods for time-dependent Lagrangian systems.Examples

However if we select the approximation using the final point we obtain the discrete Lagrangian

$$L_{d,h}^{k,\text{end}}(x_k, x_{k+1}) = hL(x_{k+1}, \frac{x_{k+1} - x_k}{h}, (k+1)h)$$

and the corresponding first-order discrete Euler-Lagrange equations for the Nesterov Lagrangian are:

$$x_{k+2} - x_{k+1} = \frac{(k+1)^3 \left(x_{k+1} - x_k - h^2 \nabla f(x_{k+1})\right)}{(k+2)^3}$$

Midpoint discretization

Another typical option is to use for the discretization of the action is to use the midpoint rule:

$$L_{d,h}^{k,\mathrm{mp}}(x_k, x_{k+1}) = hL(\frac{x_k + x_{k+1}}{2}, \frac{x_{k+1} - x_k}{h}, kh + \frac{h}{2})$$

In the case of the Nesterov Lagrangian the method is second order in h although the discrete equations are implicit:

$$\begin{aligned} 0 = & -(2k+3)^3 \left(2\left(\frac{x_{k+2} - x_{k+1}}{h}\right) + h\nabla f\left(\frac{x_{k+1} + x_{k+2}}{2}\right)\right) \\ & + (2k+1)^3 \left(2\left(\frac{x_{k+1} - x_k}{h}\right) - h\nabla f\left(\frac{x_{k-1} + x_k}{2}\right)\right) \end{aligned}$$

A Störmer–Verlet method for Brergman Lagrangian systems

$$L(x, v, t) = \frac{1}{2} e^{-\alpha(t)+\gamma(t)} \|v\|^2 - e^{\alpha(t)+\gamma(t)+\beta(t)} f(x)$$

and assuming the ideal scaling conditions by Wibosono et al (2016), that is,

$$\alpha(t) = \ln \mathbf{p} - \ln t \ , \qquad \beta(t) = \mathbf{p} \ln t + \ln C, \qquad \gamma(t) = \mathbf{p} \ln t$$

then we can write the Lagrangian as

$$L(x, v, t) = \frac{1}{2\mathbf{p}} t^{\mathbf{p}+1} \|v\|^2 - C\mathbf{p} t^{2\mathbf{p}-1} f(x)$$

If $\mathbf{p} = 2$ and $C = 1/4$

A Störmer–Verlet method for Brergman Lagrangian systems

Taking

$$L_{d,h}^{k,SV}(x_k, x_{k+1}) = \frac{h}{4\mathbf{p}}\left[(kh)^{\mathbf{p}+1} + ((k+1)h)^{\mathbf{p}+1}\right] \| \frac{q_{k+1} - q_k}{h} \|^2$$
$$- \frac{h}{2} C\mathbf{p}\left[(kh)^{2\mathbf{p}-1} f(x_k) + ((k+1)h)^{2\mathbf{p}-1} f(x_{k+1})\right]$$

A Störmer–Verlet method for Brergman Lagrangian systems

Denoting $p_{k+1/2} = \frac{1}{2\mathbf{p}}\left[(kh)^{\mathbf{p}+1} + ((k+1)h)^{\mathbf{p}+1}\right](x_{k+1} - x_k)/h$, the previous equations are rewritten in the form

$$
\begin{aligned}
p_{k+1/2} &= p_k - \frac{h}{2} C\mathbf{p}(kh)^{2\mathbf{p}-1}\nabla f(x_k),\\
x_{k+1} &= x_k + \frac{2\mathbf{p}h}{[(kh)^{\mathbf{p}+1} + ((k+1)h)^{\mathbf{p}+1}]} p_{k+1/2},\\
p_{k+1} &= p_{k+1/2} - \frac{h}{2} C\mathbf{p}((k+1)h)^{2\mathbf{p}-1}\nabla f(x_{k+1})
\end{aligned}
$$

Simple Test Function

Rosenbrock function, 1960

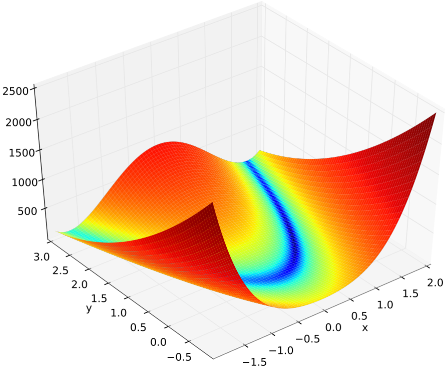
$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$

Steep well, flat valley

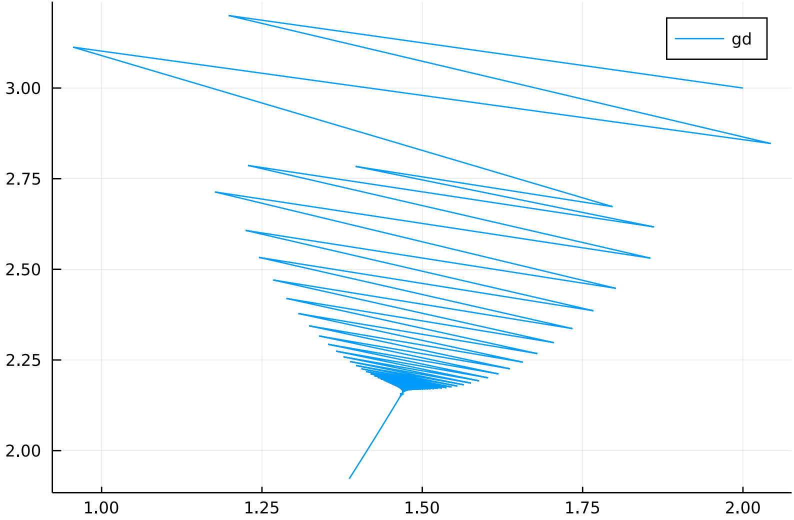
Banana shaped

Global minimum at (a, a^2)

$$f(a, a^2) = 0$$

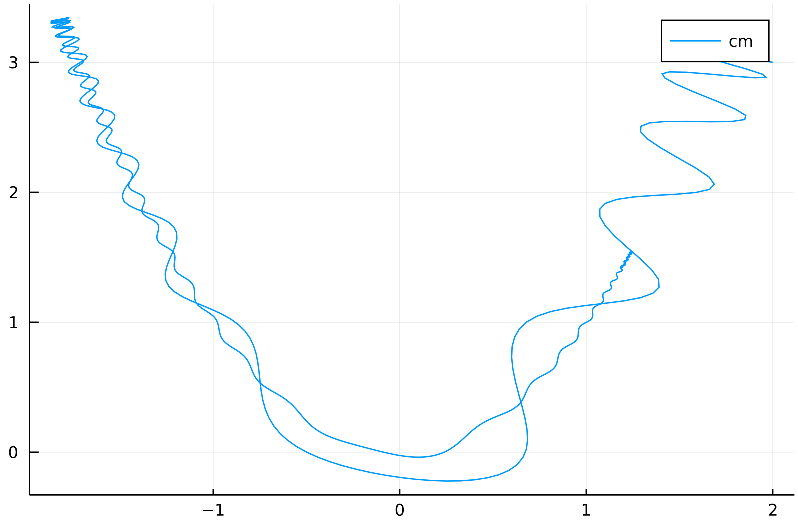


Gradient descent



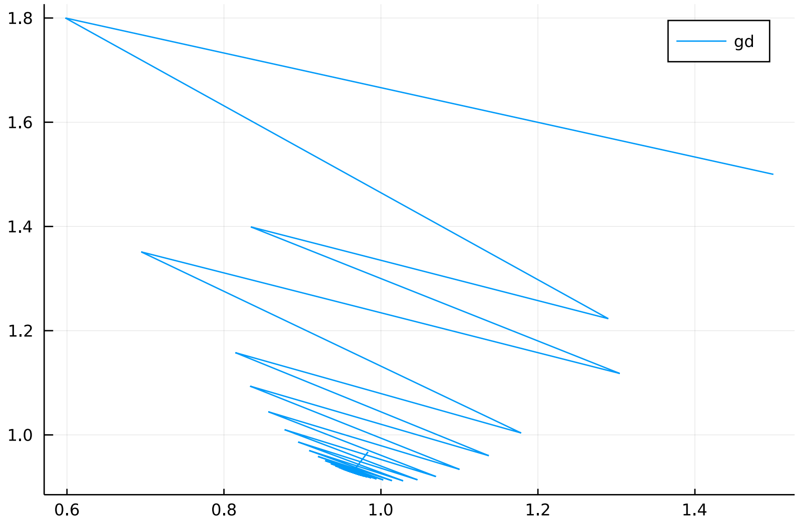
[][]

CM



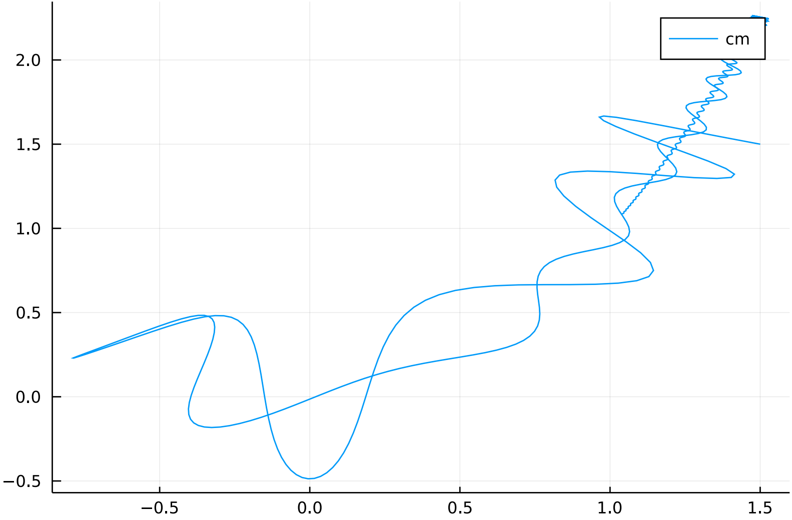
[] []

Gradient descent



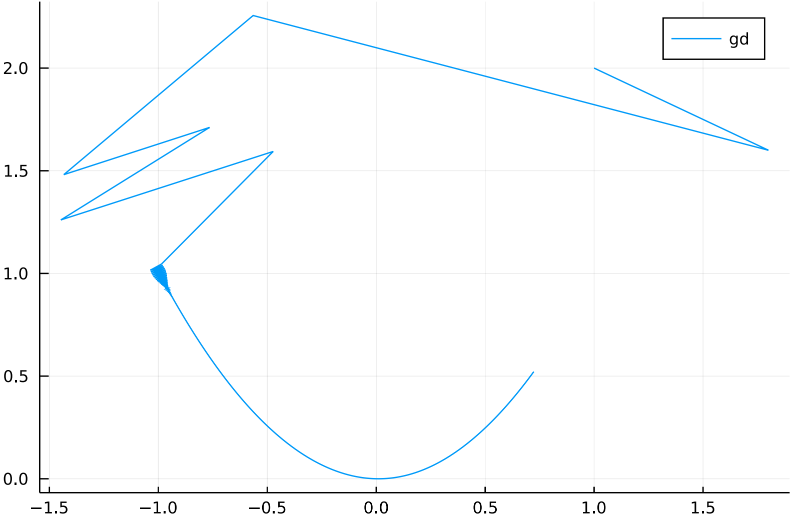
[][]

CM



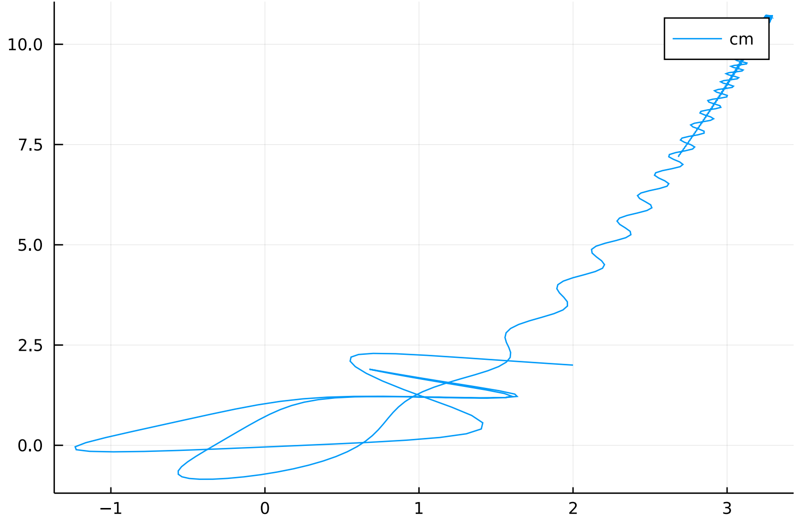
[][]

Gradient descent



[][]

CM



[][]

Discrete Lagrange-d'Alembert principle

Now, our intention is to continue looking for numerical approximations to the Euler-Lagrange equations given by a Bregman Lagrangian but additionally adding an external force that decreases jointly with the h parameter. With it we will obtain new algorithms whose behavior resembles that of the Nesterov method.

Discrete Lagrange-d'Alembert principle

Now, our intention is to continue looking for numerical approximations to the Euler-Lagrange equations given by a Bregman Lagrangian but additionally adding an external force that decreases jointly with the h parameter. With it we will obtain new algorithms whose behavior resembles that of the Nesterov method.

Fortunately, discrete mechanics is also adapted to the case of external forces. To this end, in addition to a time-dependent Lagrangian function $L\colon TQ \times \mathbb{R} \to \mathbb{R}$ we have an external force given by a fibre preserving mapping $f\colon TQ \times \mathbb{R} \to T^*Q$ given locally by

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

Discrete Lagrange-d'Alembert principle

$$\delta \int_0^h L(x(t), \dot{x}(t), t)\, dt + \int_0^h F(x(t), \dot{x}(t), t) \delta x(t)\, dt = 0 \ ,$$

for all $\delta x(t) \in T_{x(t)} Q$.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) - \frac{\partial L}{\partial x^i} = F_i$$

Discrete Lagrange-d'Alembert principle

To discretize these equations we consider as before a family of Lagrangian functions $L_d^k\colon Q \times Q \to \mathbb{R}$ and Discrete Lagrange-d'Alembert principle two discrete forces $(F_d^k)^+, (F_d^k)^-\colon Q \times Q \to T^*Q$, which are fibre preserving in the sense that $\pi_Q \circ (F_d^k)^{\mp} = pr_{\mp}$ where $pr_-(x, x') = x$ and $pr_+(x, x') = x'$. Aa discrete version of the Lagrange-d'Alembert principle for the discrete forced system given by L_d^k and F_d^k:

$$
\begin{aligned}
0 &= \delta \sum_{k=0}^{N-1} L_d^k(x_k, x_{k+1}) + \sum_{k=0}^{N-1} \langle F_d^k(x_k, x_{k+1}), (\delta x_k, \delta x_{k+1}) \rangle \\
&= \delta \sum_{k=0}^{N-1} L_d^k(x_k, x_{k+1}) + \sum_{k=0}^{N-1} \left[(F_d^k)^-(x_k, x_{k+1})\delta x_k + (F_d^k)^+(x_k, x_{k+1})\delta x_{k+1} \right]
\end{aligned}
$$

for all variations $\{\delta q_k\}_{k=0}^N$ vanishing at the endpoints, that is, $\delta q_0 = \delta q_N = 0$.

Discrete Lagrange-d'Alembert principle

This is equivalent to the forced discrete Euler-Lagrange equations:

$$
\begin{gathered}
D_1 L_d^{k+1}(x_{k+1}, x_{k+2}) + D_2 L_d^k(x_k, x_{k+1}) \\
+(F_d^{k+1})^-(x_{k+1}, x_{k+2}) + (F_d^k)^+(x_k, x_{k+1}) = 0
\end{gathered}
$$

Lemma

Given $f: Q \to \mathbb{R}$, consider the SODE

$$\ddot{x} + \nu(t)\dot{x} + \eta(t)\nabla f(x) = \varepsilon\Big[\eta(t)\nabla f(x)\Big]\,, \tag{EL}$$

where $\nu, \eta\colon \mathbb{R}_+ \to \mathbb{R}$ and $\varepsilon \in \mathbb{R}$. Then (EL) corresponds to the equation of motion of the time dependent Lagrangian system

$$L(x,\dot{x},t) = a(t)\tfrac{1}{2}\|\dot{x}\|^2 - b(t)f(x)\,,$$
$$F(x,\dot{x},t) = \varepsilon a(t)\left[\frac{b(t)}{a(t)}\nabla f(x)\right]\,,$$

where

$$a(t) = \exp(\textstyle\int_0^t \nu(s))\,, b(t) = a(t)\eta(t)\,.$$

Theorem (C. M. Campos, DMdD, A Mahillo, JMLR (2022))

Given $f : Q \to \mathbb{R}$, consider the time dependent discrete Lagrangian system

$$\begin{aligned} L_d(z_0, z_1, k) &= a_k \tfrac{1}{2} \left\| z_1 - z_0 \right\|^2 - b_k^- f(z_0) - b_{k+1}^+ f(z_1)\,, \\ F_d^-(z_0, z_1, k) &= \tfrac{a_{k-1}}{a_k} (b_k^- + b_k^+) \nabla f(z_0)\,, \ \textit{and} \\ F_d^+(z_0, z_1, k) &= \quad -(b_k^- + b_k^+) \nabla f(z_0)\,. \end{aligned}$$

where $\{a_k\}_{k\geq 0}$, $\{b_k^\pm\}_{k\geq 0}$, are arbitrary sequences. If a_k is never null, then the free and forced equations of motion for L_d and (L_d, F_d^-, F_d^+) are

$$\begin{aligned} y_{k+1} &= x_k - \eta_k \nabla f(x_k) & \qquad \bar{y}_{k+1} &= \bar{x}_k - \eta_k \nabla f(\bar{x}_k) \\ x_{k+1} &= y_{k+1} + \mu_k (x_k - x_{k-1}) & \qquad \bar{x}_{k+1} &= \bar{y}_{k+1} + \mu_k (\bar{y}_{k+1} - \bar{y}_k) \end{aligned}$$

where

$$\mu_{k+1} = \frac{a_k}{a_{k+1}} \eta_k = \frac{b_k^- + b_k^+}{a_k}\,.$$

And conversely...

Simple Test Function

Rosenbrock function, 1960

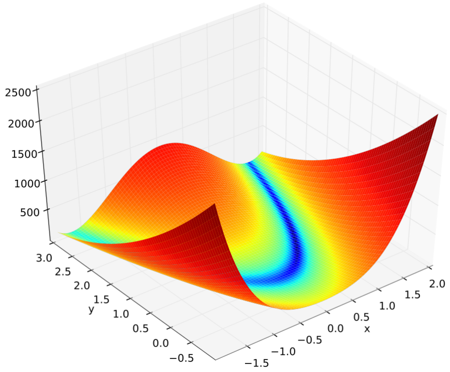
$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$

Steep well, flat valley

Banana shaped

Global minimum at (a, a^2)

$$f(a, a^2) = 0$$

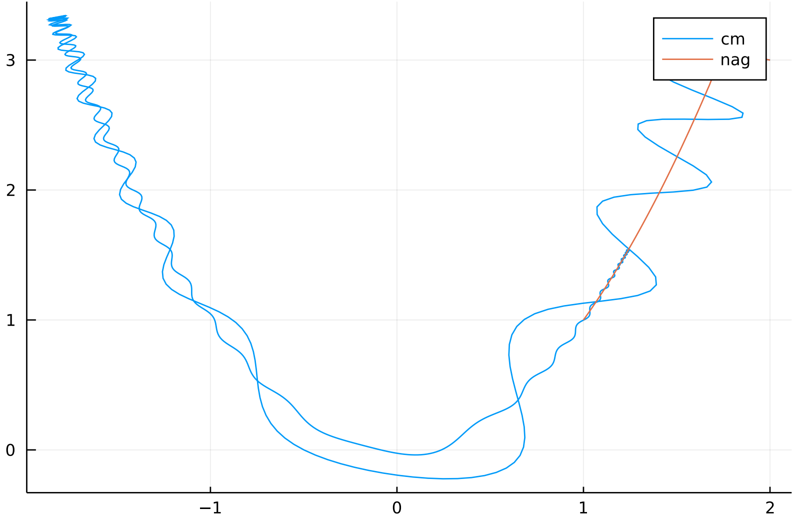


The YATF function

$$g(x, y) = \sin\left(\tfrac{1}{2}x^2 - \tfrac{1}{4}y^2 + 3\right)\cos\left(2x + 1 - e^y\right)$$

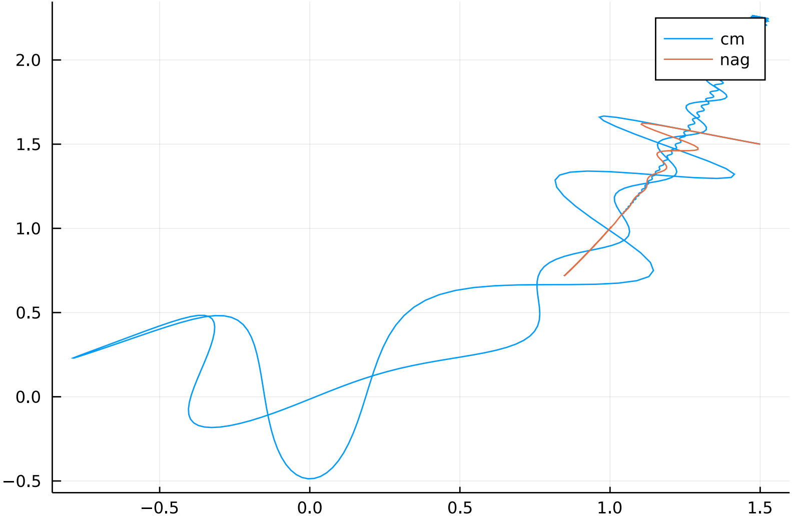
Yet Another Test Function

CM versus NAG



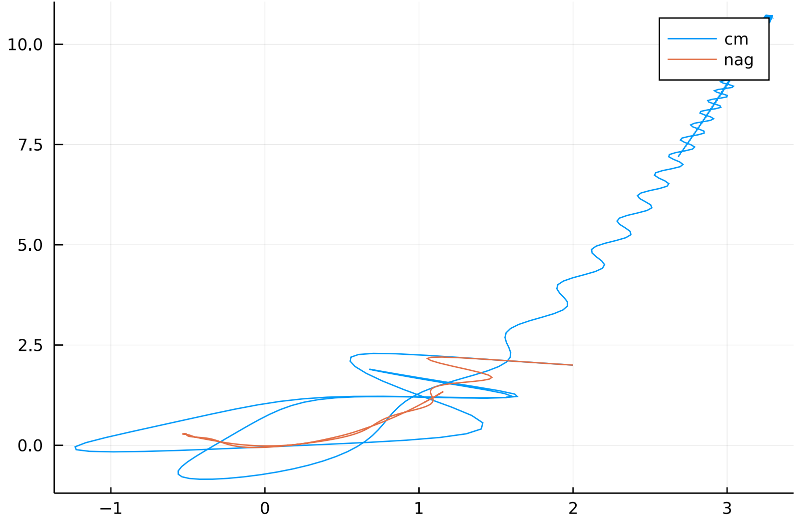
[][]

CM versus NAG



[][]

CM versus NAG



[][]

Thanks a lot!!!