Tutorial: Differential Variational Inequalities and Mechanical Contact Problems

David Stewart

Department of Mathematics

University of Iowa

BIRS Meeting on
Simulation of Non-Smooth Mechanical Systems

Collaborators & contributors

Jeongho Ahn
Mihai Anitescu
Patrick Ballard
Bernard Brogliato
Marius Cocou
Jiří Jarušek
João A. C. Martins
Manuel Monteiro-Marques
Jong-Shi Pang
Florian Potra
Michelle Schatzman
Jeffrey Trinkle
Theodore Wendt

What is a DVI?

A **Differential Variational Inequality (DVI)** has a differential part

$$\frac{dx}{dt} = f(t, x) + B(t, x)\, z(t), \qquad x(t_0) = x_0$$

and a variational inequality (VI) part

$$\begin{array}{lll} K & & \text{a closed and convex set} \\ z(t) \in K & \& & 0 \leq \langle G(t, x, z),\, \widetilde{z} - z(t) \rangle \qquad \text{for all } \widetilde{z} \in K. \end{array}$$

If K is also a cone ($x \in K$ & $\alpha \geq 0$ implies $\alpha x \in K$) then the variational inequality part becomes **complementarity problem (CP)**

$$\begin{aligned} K \ni z(t) \quad &\perp \quad G(t, x(t), z(t)) \in K^* \qquad \text{for all } t \\ K^* \quad &= \quad \{ \, w \mid \langle w, \, v \rangle \geq 0 \text{ for all } v \in K \, \} \, . \end{aligned}$$

If $K = \mathbb{R}^n_+$ then $K^* = \mathbb{R}^n_+$ and we want

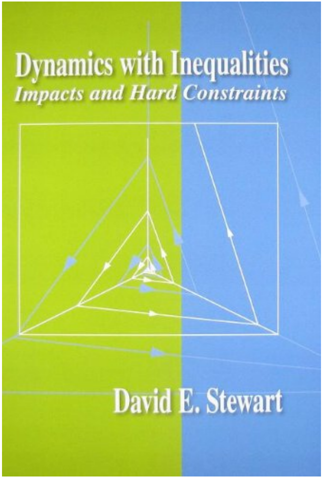
$$0 \leq z(t) \;\perp\; G(t, x(t), z(t)) \geq 0.$$

If $K = \mathbb{R}^n$ then $K^* = \{0\} \subset \mathbb{R}^n$ so $G(t, x(t), z(t)) = 0$ and $z(t)$ free in $\mathbb{R}^n$.

Why?

$$\begin{aligned} \frac{dx}{dt} &= f(t,x) + B(t,x)\,z(t), \qquad x(t_0) = x_0 \\ z(t) \in K \quad \& \quad 0 &\le \langle G(t,x,z),\, \widetilde{z} - z(t) \rangle \qquad \text{for all } \widetilde{z} \in K. \end{aligned}$$

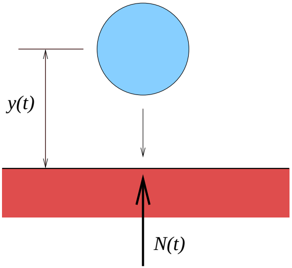
- The VI variable $z(t)$ can jump to describe discontinuous right-hand side, or even contain impulses for impact problems.
- There is a great deal of theory for variational inequalities.
- Can be applied to finite dimensional and infinite dimensional (PDE) problems.
- Goes beyond the theory of maximal monotone operator differential equations (see Brézis).



SIAM Publ., 2011

Applications

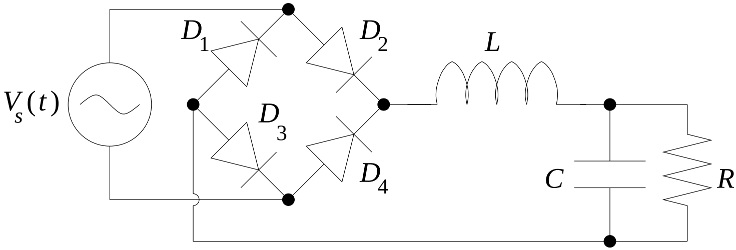
- Impact problems: (index two)

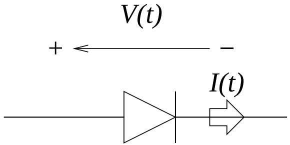


$$m\frac{d^2y}{dt^2} = -mg + N(t)$$
$$0 \leq N(t) \perp y(t) - r \geq 0$$

And PDE models as well!

Diode circuits: (index one)

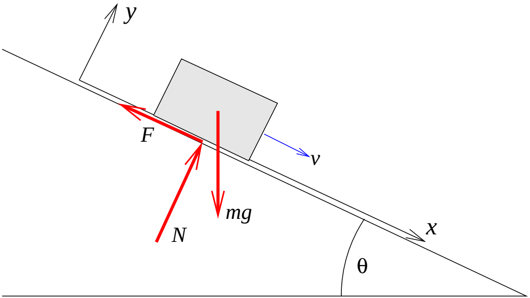




current flow

$$0 \leq -V(t) \ \perp \ I(t) \geq 0$$

- Coulomb friction: (index one)



$$
\begin{aligned}
m\frac{dv}{dt} &= mg\sin\theta - F, \\
F \in K \quad \& \quad 0 &\le v(\widetilde{F} - F) \qquad \text{for all } \widetilde{F} \in K \\
K &= mg\cos\theta[-1, +1].
\end{aligned}
$$

Index

In analogy with the concept of index for differential algebraic equations (DAE's), differential variational inequalities also have an index:

$$\begin{aligned} \frac{dx}{dt} &= f(t,x) + B(t,x)\, z(t), \qquad x(t_0) = x_0 \\ z(t) \in K \quad &\& \quad 0 \le \langle G(t,x,z), \tilde{z} - z(t) \rangle \qquad \text{for all } \tilde{z} \in K. \end{aligned}$$

The index of this DVI is the minimal m where the equation

$$\frac{d^m}{dt^m} G(t, x(t), z(t)) = b(t)$$

can be solved for $z(t)$ (after doing all substitutions of derivatives from the differential equation) as a function of t and $x(t)$.

The index of a DVI is crucial for understanding the existence and uniqueness theory.

Index zero: $G(t, x, z) = b(t)$ can be solved for z as a function of t and x. Then as long as $z \mapsto G(t, x, z)$ is a strictly monotone the VI has a unique solution that depends in a locally Lipschitz way on t and x. Substituting gives a locally Lipschitz ODE for $x(t)$. Usual theory applies.

Index one: Typically $G(t, x, z) = G(t, x)$. Then we need $\nabla_x G(t, x(t))\, B(x(t))$ to be positive definite (or elliptic in infinite dimensions) to get existence and (in addition) symmetric for uniqueness.

Index two: This occurs for many impact problems. Existence can usually be shown, but not uniqueness. Lack of uniqueness can occur even with the "coefficient of restitution" specified.

Index three and higher: Existence usually fails. There are some alternative formulations, but these usually do not satisfy the principle "limits of solutions are also solutions".

Fractional index: This makes sense for convolution complementarity problems: Find $z(t)$ given $m(t)$, $q(t)$ and K where

$$\begin{aligned} K \ni z(t) \quad &\perp \quad (m * z)(t) + q(t) \in K^* \\ (m * z)(t) \quad &= \quad \int_0^t m(t-\tau)\, z(\tau)\, d\tau. \end{aligned}$$

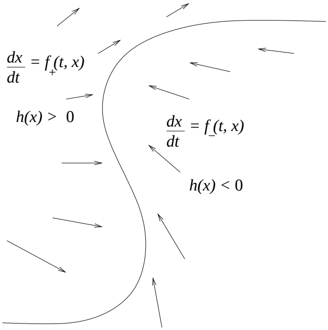
Existence and uniqueness for index $\alpha \in (0,1)$ or $\alpha = 1$ and $m(0^+)$ symmetric; existence for $\alpha \in (1,2]$.

Index one

Applications: Coulomb friction, diode circuits, traffic flow, networks of queues

Infinite dimensional versions: maximal monotone operator differential equations, parabolic variational inequalities

Also, this can handle large classes of structured Filippov discontinuous ODEs.



$$\frac{dx}{dt} = (1 - z(t))\, f_+(t,x) + z(t)\, f_-(t,x)$$
$$z(t) \in [0,1] \quad \& \quad 0 \le h(x)\,(\widetilde{z} - z(t)) \qquad \text{for all } \widetilde{z} \in [0,1]$$

Note: $B(t,x) = f_-(t,x) - f_+(t,x)$ and $G(t,x) = h(x)$ so we need to look at

$$\nabla_x G(t,x) B(t,x) = \nabla h(x)^T (f_-(t,x) - f_+(t,x))$$

If this is positive, then solutions exist and are unique; if negative then we can show solutions exist but are not unique.

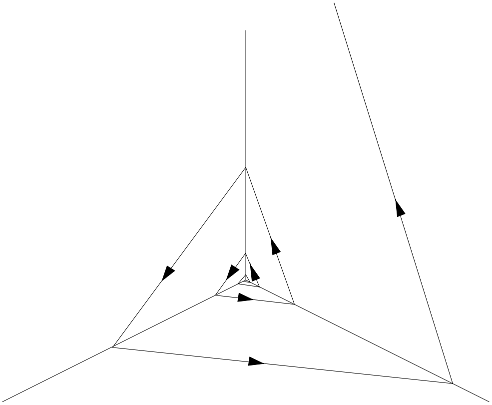
Nonuniqueness

Nonuniqueness can arise for index one DVI's when $\nabla_x G(t,x) B(t,x)$ is not positive definite: the variational inequality can have more than one solution.

The variational inequality might have unique solutions, but uniqueness for the DVI can fail. This involves ***reverse Zeno*** solutions!

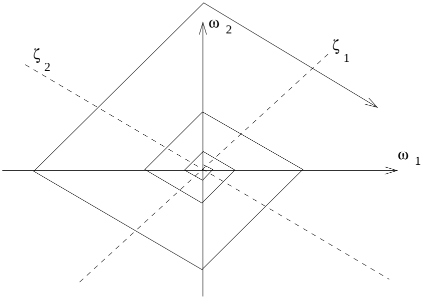
Example:

$$\begin{aligned} \frac{dw}{dt} &= M\,z(t) + f(t), \qquad w(0) = 0, \\ 0 \le z(t) &\perp w(t) \ge 0 \quad \text{for all } t, \text{ with} \\ M &= \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 3 \\ 3 & 0 & 1 \end{bmatrix}, \qquad f(t) \equiv 0. \end{aligned}$$



$w(t)$

Avi Mandelbaum has a more sophisticated example with $f(t) \not\equiv 0$ but with M 2×2 and positive definite. The function $f(t)$ cannot be analytic but can be C^∞ in his example.



$$
\begin{aligned}
\frac{d\omega}{dt} &= M\zeta(t), \\
\omega(t)^T\zeta(t) &\leq 0 \qquad \text{for all } t.
\end{aligned}
$$

Differentiation lemmas

If

$$K \ni a(t) \ \perp \ b(t) \in K^* \qquad \text{for all } t,$$

and a and b are sufficiently regular, then

$$\begin{aligned} 0 &= \langle a'(t), b(t) \rangle, \\ 0 &\geq \langle a'(t), b'(t) \rangle, \\ 0 &\leq \langle a''(t), b(t) \rangle. \end{aligned}$$

"Sufficiently regular" for the 1st case can mean

$$\begin{aligned} a &\in W^{1,p}(0,T;X), & b &\in L^q(0,T;X'), & p^{-1}+q^{-1} = 1, \\ a &\in C^1(0,T;X), & b &\in \mathcal{M}([0,T];X'), & \text{or} \\ a &\in H^{1+\alpha}(0,T;X), & b &\in H^{-\alpha}(0,T;X'). \end{aligned}$$

Energy balance

For a rigid or semi-rigid body:

$$
\begin{aligned}
M(q)\frac{dv}{dt} &= k(q,v) - \nabla V(q) + \nabla\varphi(q)\,N(t), \\
\frac{dq}{dt} &= v, \\
0 \le N(t) &\perp \varphi(q(t)) \ge 0 \qquad \text{for all } t.
\end{aligned}
$$

What happens to the energy $E(t) = \frac{1}{2}v^T M(q)v + V(q)$?

$$
\begin{aligned}
\frac{dE}{dt} &= \frac{d}{dt}\left[\frac{1}{2}v^T M(q) v + V(q)\right] \\
&= v^T M(q)\frac{dv}{dt} + \frac{1}{2}v^T \frac{d}{dt}\left(M(q)\right) v + \nabla V(q)^T v \\
&= v^T \nabla\varphi(q)\, N = \frac{d}{dt}\left(\varphi(q)\right)\, N.
\end{aligned}
$$

We know that

$$
0 \le N(t) \perp \varphi(q(t)) \ge 0 \qquad \text{for all } t.
$$

Provided $\varphi(q(t))$ and $N(t)$ are ***suitably regular*** we can use a differentiation lemma to get

$$
\frac{d}{dt}\left(\varphi(q)\right)\, N = 0.
$$

PDE & ∞-dimensional problems

Parabolic variational inequalities (PVI's):

Example: Obstacle problem:

$$u \in K \qquad \& \qquad 0 \leq \left\langle \frac{\partial u}{\partial t} - \nabla^2 u - f(t),\, v - u \right\rangle \qquad \text{for all } v \in K,$$

$$K = \left\{ w \mid w \in H^1(\Omega), \quad w \geq \varphi \right\}.$$

Existence and uniqueness follow from theory of maximal monotone operators.

Elastic & viscoelastioc bodies

Frictionless contact: $\mathbf{u}(t,\mathbf{x})$ = displacement, $\sigma(t,\mathbf{x})$ = stress tensor, $\varphi(\mathbf{x})$ = gap function (on boundary), ρ = density

$$K = \left\{ \mathbf{w} \in H^1(\Omega) \mid \varphi - \mathbf{n}\cdot\mathbf{w} \geq 0 \text{ on } \partial\Omega \right\},$$

$$K \ni \mathbf{u}(t) \quad \& \quad 0 \leq \left\langle \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - \operatorname{div}\sigma - \mathbf{f},\ \widetilde{\mathbf{u}} - \mathbf{u} \right\rangle \quad \text{for all } \widetilde{\mathbf{u}} \in K.$$

Alternative formulation using complementarity with normal contact forces $N(t,\mathbf{x})$, $\mathbf{x} \in \partial\Omega$:

$$\begin{aligned} \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} &= \operatorname{div}\sigma + \mathbf{f} && \text{in } \Omega, \\ \sigma\mathbf{n} &= -N\mathbf{n} && \text{on } \partial\Omega, \\ 0 \leq N(t,\mathbf{x}) &\perp \varphi(\mathbf{x}) - \mathbf{n}(\mathbf{x})\cdot\mathbf{u}(t,\mathbf{x}) \geq 0 && \text{for all } \mathbf{x} \in \partial\Omega. \end{aligned}$$

Stress tensor for linearized (visco-)elasticity

$$\begin{aligned}
\sigma_{ij} &= \sum_{k,l} \left(a_{ijkl}\varepsilon_{kl} + b_{ijkl}\frac{\partial \varepsilon_{kl}}{\partial t} \right), \\
\varepsilon_{kl} &= \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right).
\end{aligned}$$

For pure elasticity $b_{ijkl} = 0$ for all i, j, k, l. Also we need the trace map $\gamma \colon H^1(\Omega)^d \to H^{1/2}(\partial\Omega)$ ($\gamma \mathbf{u} = -\mathbf{n}\cdot\mathbf{u}|\partial\Omega$) and its adjoint $\gamma^* \colon H^{-1/2}(\partial\Omega) \to H^1(\Omega)'^d$:

$$\begin{aligned}
\rho\frac{\partial^2 \mathbf{u}}{\partial t^2} &= -\mathcal{A}\mathbf{u} - \mathcal{B}\frac{\partial \mathbf{u}}{\partial t} + \mathbf{f} + \gamma^* N, \\
0 \le N &\perp \gamma\mathbf{u} + \varphi \ge 0.
\end{aligned}$$

$$
\begin{aligned}
\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} &= -\mathcal{A}\mathbf{u} - \mathcal{B}\frac{\partial \mathbf{u}}{\partial t} + \mathbf{f} + \gamma^* N, \\
0 \leq N &\perp \gamma \mathbf{u} + \varphi \geq 0.
\end{aligned}
$$

Existence for $\mathcal{B}$ elliptic shown by Marius Cocu; extended by Cocu, and by Kuttler & Shillor.

Uniqueness is open.

Energy balance mostly complete (based on differentiation lemma).

Existence for $\mathcal{B} = 0$ for the wave equations ($\mathbf{u}(t, \mathbf{x}) \in \mathbb{R}$, $\mathcal{A} = -\Delta$) via compensated compactness.

Existence for $\mathcal{B} = 0$ for (even) isotropic elasticity unknown.

Numerical methods

Index two: rigid-body dynamics

The methods are somewhat specialized, but typically low order in order to ensure stability and positivity. The methods must be implicit, at least for the complementarity conditions and/or variational inequality conditions.

Continuous-time system:

$$
\begin{aligned}
M(q)\frac{dv}{dt} &= n(q)\,N + D(q)\,\beta - \nabla V(q) + k(q,v) + F_{ext}(t), \\
\frac{dq}{dt} &= v, \\
0 &\leq N \quad \perp \quad f(q) \geq 0, \\
0 &\in \mu\, D(q)^T v^+ + \lambda\, \partial\psi(\beta), \\
0 &\leq \lambda \quad \perp \quad \mu N - \psi(\beta) \geq 0, \\
0 &= n(q(t))^T(\varepsilon v^-(t) + v^+(t)) \quad \text{if} \quad f(q(t)) = 0.
\end{aligned}
$$

Time-discretization:

$$
\begin{aligned}
M(q^{l+1})(v^{l+1} - v^l) &= n(q^l) c_n^{l+1} + \widetilde{D}(q^l)\widetilde{\beta}^{l+1} \\
&\quad + h[-\nabla V(q^l) + k(q^l, v^l) + F_{ext}(t_l)] \\
q^{l+1} - q^l &= h\, v^{l+1}, \\
0 \le c_n^{l+1} &\perp n(q^l)^T (v^{l+1} + \varepsilon v^l) \ge 0, \\
0 \le \widetilde{\beta}^{l+1} &\perp \lambda^{l+1} e + \widetilde{D}(q^l)^T v^{l+1} \ge 0, \\
0 \le \lambda^{l+1} &\perp \mu\, c_n^{l+1} - e^T \widetilde{\beta}^{l+1} \ge 0.
\end{aligned}
$$

Index-one DVIs

Here we can use more sophisticated methods, such as high-order implicit Runge–Kutta methods:

Exact problem:

$$\begin{aligned} \frac{dx}{dt}(t) &= f(x(t)) + B(x(t))\, z(t), \qquad x(t_0) = x_0, \\ z(t) \in K \quad \& \quad 0 &\leq \langle \widetilde{z} - z(t),\, G(x(t)) \rangle \qquad \text{for all } \widetilde{z} \in K. \end{aligned}$$

Time discretization:

$$\begin{aligned} v_{ni} &= x_n + h\sum_{j=1}^{s} a_{ij}\,\left[f(v_{nj}) + B(v_{nj})z_{nj}\right], \\ z_{ni}\in K \quad \& \quad 0 &\le \langle \widetilde{z}_{ni} - z_{ni},\, G(v_{ni})\rangle \qquad \text{for all } \widetilde{z}_{ni}\in K, \\ x_{n+1} &= x_n + h\sum_{j=1}^{s} b_j\,\left[f(v_{nj}) + B(v_{nj})z_{nj}\right], \end{aligned}$$

based on the Butcher tableau

$$\begin{array}{c|c} \mathbf{c} & A \\ \hline & \mathbf{b}^T \end{array} \qquad \text{or} \qquad \begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array}$$

Conditions for Runge–Kutta methods

The method needs to satisfy the following conditions:

1. it is algebraically stable (that is, $\mathbf{b} \geq 0$ and $\operatorname{diag}(\mathbf{b})\, A + A^T \operatorname{diag}(\mathbf{b}) - \mathbf{b}\mathbf{b}^T$ is positive semi-definite) ,
2. it satisfies conditions $\mathrm{B}(p)$ and $\mathrm{C}(q)$ of Butcher,
3. there exists a diagonal matrix D with positive diagonal entries such that $DA + A^T D$ is (symmetric) positive definite, and
4. the method is *stiffly accurate*; that is, $\mathbf{b}^T = \mathbf{e}_s^T A$.

Then the method has order $\mathcal{O}(h^q)$. Radau IIA methods satisfy these conditions (with $q = s$).