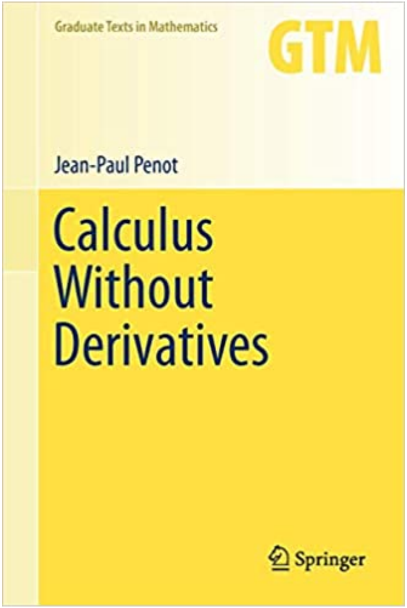
One World Optimization Seminar
April 19, 2021

Sensitivity Analysis without Derivatives

Asen L. Dontchev

The University of Michigan

Supported by AFOSR Award FA9550-20-1-0385


Graduate Texts in Mathematics
GTM
Jean-Paul Penot
Calculus Without Derivatives
Springer

Outline of the talk

1. Introduction
2. Fiacco-Mccormick theorem
3. Hildebrand-Graves theorem
4. Robinson's theorem
5. Strong regularity
6. Application to nonlinear programming
7. Applications to optimal control

– strong regularity of the optimality system
– uniform convergence of Newton/SQP method
– existence of optimal feedback
– model predictive control

Sensitivity Analysis: Estimating the effect of changes (perturbations, approximations, errors, inaccuracies) of a (optimization/control) problem on solutions.

Has been traditionally identified with determining derivatives of the optimal value and/or an optimal solution with respect to parameters.

But for optimization problems with constraints finding derivatives might be challenging (generalized derivatives).

And even more difficult for optimal control problems, where in addition one has to differentiate in infinite dimensions.

Sensitivity analysis has direct applications to proving convergence of optimization algorithms.

A. V. Fiacco, G. P. Mccormick, Nonlinear programming: sequential unconstrained minimization techniques, Wiley 1968.

$$\text{minimize } \; g_0(p, x)$$

subject to

$$g_i(p, x) \leq 0, \quad i = 1, \ldots, s$$
$$x \in \boldsymbol{R}^n, \quad p \in \boldsymbol{R}^k$$

Lagrange function

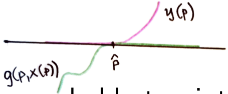
$$L(p, x, y) = g_0(p, x) + \sum_{i=1}^{s} y_i g_i(p, x)$$

KKT conditions (under a constraint qualification)

....

Theorem (Fiacco–Mccormick).

Let $\bar{x}$ be a solution for $\bar{p}$ and $\bar{y}$ be an associated Lagrange multiplier. Let $g_i, i = 0, \ldots, m$ be twice continuously differentiable around $(\bar{p}, \bar{x})$. Suppose that the following conditions are satisfied at $(\bar{p}, \bar{x}, \bar{y})$:
(i) the gradients of $D_x g_i(\bar{p}, \bar{x})$ of the active constraints are linearly independent;
(ii) the second-order sufficiency;
(iii) the strict complementarity slackness.
Then the mapping $p \mapsto (\bar{x}(p), \bar{y}(p))$ has a continuously differentiable single-valued graphical localization around $\bar{p}$ for $(\bar{x}, \bar{y})$.



Condition (iii) may never hold at points where a constraint change from active to nonactive.

Hildebrand-Graves Theorem (slightly updated)

Lipschitz modulus

$$\operatorname{lip}(f;\bar{x}) := \limsup_{\substack{x',x\to\bar{x},\\ x\neq x'}} \frac{\|f(x')-f(x)\|}{\|x'-x\|}.$$

Theorem (T. Hildebrand, L. M. Graves, TAMS 29 (1927) 127–153).

Let X be a Banach space and consider a function $f : X \to X$ and a linear bounded mapping $A : X \to X$ which is invertible. Suppose that

$$\operatorname{lip}(f - A; \bar{x}) \cdot \|A^{-1}\| < 1.$$

Then f^{-1} has a single-valued localization around $f(\bar{x})$ for $\bar{x}$ which is Lipschitz continuous.

The H-G IFT implies the classical (Dini) IFT: f is strictly differentiable at $\bar{x} \iff \operatorname{lip}(f - Df(\bar{x}); \bar{x}) = 0.$

T. Hildebrand (1888–1980) and L. Graves (1886–1973)

Robinson's Theorem

S. M. Robinson, Strongly regular generalized equations, Math. of OR 5 (1980) 43–62.

This is an implicit function theorem for variational inequalities

$$(*) \qquad f(p,x) + N_C(x) \ni 0,$$

where $f : P \times X \to X^*$, P and X are Banach spaces, and N_C is the normal cone mapping to the convex and closed set $C \subset X$:

$$N_U(x) = \begin{cases} \{y \in X^* \mid \langle y, v - x\rangle \le 0 \text{ for all } v \in U\} & \text{if } x \in C, \\ \emptyset & \text{otherwise.} \end{cases}$$

Robinson's theorem, slightly simplified

Theorem.

Let $\bar{x}$ is a solution of $(*)$ for $\bar{p}$ and assume that the inverse of the linearized mapping

$$y \mapsto (f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(\cdot - \bar{x}) + N_C(\cdot))^{-1}(y)$$

has a Lipschitz continuous single-valued localization at 0 for $\bar{x}$. Then the solution mapping of $(*)$ has a Lipschitz continuous single-valued localization at $\bar{p}$ for $\bar{x}$.

The property in blue was called by Robinson strong regularity.

Addition: If the localization of the inverse is (directionally, semi-, Fréchet, strictly) differentiable then so is the localization of the solution mapping.

For $F = 0$ we (almost) obtain the classical implicit function

Strong regularity of generalized equations

Strong regularity of F at $\bar{x}$ for $\bar{y}$: $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and F^{-1} has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$.

Generalized equations

$$f(p,x) + F(x) \ni 0,$$

where $f : P \times X \to Y$ is a function and F is a set-valued mapping.

Solution mapping

$$p \mapsto S(p) = \left\{ x \,\middle|\, f(p,x) + F(x) \ni 0 \right\}$$

Robinson's theorem, extended

Theorem.

Let X be a complete metric space, let Y be a linear metric space with a shift-invariant metric and let P be a metric space. For a function $f : P \times X \to Y$ and a set-valued mapping $F : X \rightrightarrows Y$, consider the generalized equation $f(p, x) + F(x) \ni 0$ with solution mapping S having $\bar{x} \in S(\bar{p})$. Let κ and μ be positive constants such that $\kappa\mu < 1$.

Let $h : X \to Y$ be such that $h(\bar{x}) = f(\bar{p}, \bar{x})$ and $\widehat{\mathrm{lip}}_x(f - h; (\bar{p}, \bar{x})) < \mu$. Also suppose that $h + F$ is strongly regular at $\bar{x}$ for 0 with a Lipschitz continuous single-valued localization around 0 for $\bar{x}$ having Lipschitz constant κ.

Then the mapping S has a single-valued localization s around $\bar{p}$ for $\bar{x}$ which is Lipschitz continuous near $\bar{p}$ with

$$\mathrm{lip}(s; \bar{p}) \leq \frac{\kappa}{1 - \kappa\mu} \widehat{\mathrm{lip}}_p(f; (\bar{p}, \bar{x})).$$

Keeping track of the constants

Theorem.

Consider a mapping $F : X \rightrightarrows Y$ with $\bar{y} \in F(\bar{x})$, where X and Y are Banach spaces, and suppose that there exist constants $\kappa \geq 0$, $a > 0$ and $b > 0$ such that the truncated inverse mapping

$$B_b(\bar{y}) \ni y \mapsto F^{-1}(y) \cap B_a(\bar{x})$$

is single-valued and Lipschitz continuous on $B_b(\bar{y})$ with Lipschits constant κ. Let $\mu > 0$ be such that $\kappa\mu < 1$ and let $\kappa' > \kappa/(1 - \kappa\mu)$. Then for every positive α and β such that

$$\alpha \leq a/2, \quad 2\mu\alpha + 2\beta \leq b \quad \text{and} \quad 2\kappa'\beta \leq \alpha$$

and every function $f : X \to Y$ satisfying $\|f(\bar{x})\| \leq \beta$ and $\|f(x) - f(x')\| \leq \mu\|x - x'\|$ for every $x, x' \in B_{2\alpha}(\bar{x})$, the mapping $y \mapsto (f + F)^{-1}(y) \cap B_\alpha(\bar{x})$ is a Lipschitz continuous function on $B_\beta(\bar{x})$ with Lipschitz constant κ'.

Strong regularity of the KKT mapping

$$\text{minimize } g_0(p,x) \text{ subject to } g_i(p,x) \begin{cases} \leq 0 & \text{for } i \in [1,s], \\ = 0 & \text{for } i \in [s+1,m], \end{cases}$$

the functions $g_0, g_1, \ldots, g_m$ are twice continuously differentiable.
Lagrangian

$$L(p,x,y) = g_0(p,x) + y_1 g_1(p,x) + \cdots + y_m g_m(p,x)$$

Under (Mangasarian-Fromovitz) constraint qualification, the KKT optimality system is

$$(1) \qquad f(p,x,y) + N_E(x,y) \ni (0,0),$$

where

$$\begin{cases} f(p,x,y) = (\nabla_x L(p,x,y), -\nabla_y L(p,x,y)), \\ E = \boldsymbol{R}^n \times [\boldsymbol{R}^s_+ \times \boldsymbol{R}^{m-s}]. \end{cases}$$

KKT solution mapping of (1)

Notation

$$I = \{ i \in [1,m] \,|\, g_i(\bar{p},\bar{x}) = 0 \},$$
$$I_0 = \{ i \in [1,s] \,|\, g_i(\bar{p},\bar{x}) = 0 \text{ and } \bar{y}_i = 0 \}$$
$$M^+ = \{ w \in \mathbb{R}^n \,|\, w \perp \nabla_x g_i(\bar{p},\bar{x}) \text{ for all } i \in I \backslash I_0 \}.$$

Theorem.

For the KKT solution mapping S, let $(\bar{x},\bar{y}) \in S(\bar{p})$ and assume that the following conditions are both fulfilled:

(a) the gradients $\nabla_x g_i(\bar{p},\bar{x})$ for $i \in I$ are linearly independent,

(b) $\langle w, \nabla^2_{xx} L(\bar{p},\bar{x},\bar{y}) w \rangle > 0$ for every nonzero $w \in M^+$.

Then the mapping S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $(\bar{x},\bar{y})$.

Moreover, if the matrix $[\nabla_p g_i(\bar{p},\bar{x})]$ has full rank, then conditions (a) and (b) are also necessary (ample parameterization).

Optimal Control

$$\text{Minimize} \quad \int_0^T \varphi(p, x(t), u(t))dt$$

subject to

$$\dot{x}(t) = g(p, x(t), u(t)), \quad x(0) = 0,$$

$$u(t) \in U \quad \text{for a.e. } t \in [0, T], \quad u \in L^\infty, \quad x \in W_0^{1,\infty}$$

$p \in \boldsymbol{R}^d$ is a parameter with reference value $\bar{p}$.

Standing Assumptions: U is closed and convex. The functions φ and g are twice differentiable and their second derivatives are locally Lipschitz continuous. There exists a locally optimal solution $(\bar{x}, \bar{u})$ of the problem for $\bar{p}$.

First-order necessary optimality condition

Hamiltonian

$$H(p, x, u, q) = \varphi(p, x, u) + q^T g(p, x, u)$$

Optimality system

$$\left\{ \begin{array}{lcl} \dot{x}(t) & = & g(p, x(t), u(t)), \quad x_0 = 0, \\ \dot{q}(t) & = & -\nabla_x H(p, x(t), u(t), q(t)), \quad q(T) = 0, \\ 0 & \in & \nabla_u H(p, x(t), u(t), q(t)) + N_U(u(t)), \end{array} \right.$$

all for a.e. $t \in [0, T]$.
Can be written as a generalized equation

$$f(p, x) + F(x) \ni 0$$

which is not a variational inequality.

Strong regularity of the optimality system

Define the matrices

$$A(t) = \nabla_x \bar{g}(t), \quad B(t) = \nabla_u \bar{g}(t),$$
$$Q(t) = \nabla_{xx} \bar{H}(t), \quad S(t) = \nabla_{xu} \bar{H}(t), \quad R(t) = \nabla_{uu} \bar{H}(t),$$

– COERCIVITY: there exists a constant $\alpha > 0$ such that

$$\int_0^T (y(t)^T Q(t) y(t) + w(t)^T R(t) w(t) + 2y(t)^T S(t) w(t)) dt$$

$$\geq \alpha \int_0^T |w(t)|^2 dt$$

for all $y \in W^{1,2}, y(0) = 0, w \in L^2$ such that
$\dot{y}(t) = A(t)y(t) + B(t)w(t), \ y(0) = 0,$ and $w(t) \in U - U$ for a.e. $t \in [0, T]$.

- ISOLATEDNESS: The optimal control $\bar{u}$ has a representative which is an isolated solution of the inclusion

$$\nabla_u H(\bar{p}, \bar{x}(t), u, \bar{q}(t)) + \mathcal{N}_U(u) \ni 0 \quad \text{for all } t \in [0, T].$$

Under coercivity and isolatedness

– the optimal control $\bar{u}$ has a representative which is Lipschitz continuous in time $t \in [0, T]$;

– the mapping describing the optimality system for $\bar{p}$ is strongly regular at $(\bar{x}, \bar{u}, \bar{q})$ for 0;

– the mapping "$p \mapsto$ the set of solutions of the optimality system" has a Lipschitz continuous single-valued localization around $\bar{p}$.

Discrete approximation

Choose uniform grid $\{t_i\}$; $t_{i+1} - t_i = h = T/N$ and consider

$$\text{Minimize} \ \sum_{i=0}^{N-1} h\varphi(p, x_i, u_i)$$

subject to

$$x_{i+1} = x_i + hg(\bar{p}, x_i, u_i), \quad u_i \in U \quad \text{for } i = 0, 1, \dots, N-1, \quad x_0 = 0,$$

Optimality system

$$\left\{ \begin{array}{lcl} x_{i+1} & = & x_i + hg(\bar{p}, x_i, u_i), \quad x_0 = 0, \\ q_{i-1} & = & q_i + h\nabla_x H(\bar{p}, x_i, u_i, q_i), \quad q_{N-1} = 0, \\ 0 & \in & \nabla_u H(\bar{p}, x_i, u_i, q_i) + N_U(u_i), \end{array} \right.$$

Under coercivity and isolatedness, if $\bar{u}^N$ is a piecewise constant interpolation of the optimal control of the discretized problem, then

$$\|\bar{u} - \bar{u}^N\|_C = O(h).$$

The proof uses (a version of) Robinson's theorem.

Newton's method

Solving equations: $f(x^k) + Df(x^k)(x^{k+1} - x^k) = 0.$

Solving generalized equations:
$f(x^k) + Df(x^k)(x^{k+1} - x^k) + F(x^{k+1}) \ni 0$

In optimization: Sequential Quadratic Programming (SQP)

N. Josephy (1979): Strong regularity implies quadratic convergence to a solution when the starting point is close to it.

Moreover, under coercivity and isolatedness, when applied to a discretized optimal control problem with a parameter, the convergence is uniform with respect to the discretization step and small changes of the parameter.

Example: Spacecraft reorientation

State: angular velocities $\omega \in \mathbb{R}^3$
Control: torques $u \in \mathbb{R}^3$, subject to box constraints
State equation

$$\begin{cases} \dot{\theta} = E(\theta)\omega \\ \dot{\omega} = \hat{\omega} J\omega + u, \end{cases}$$

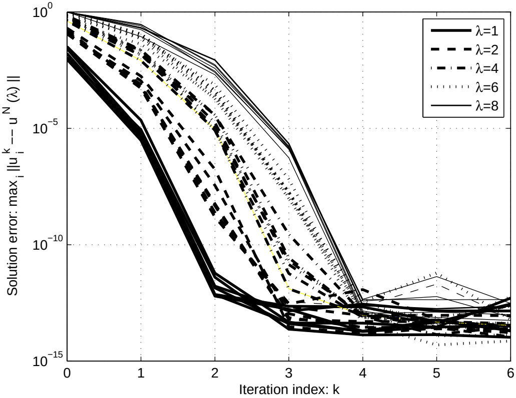
where

$$E(\theta) = \frac{1}{\cos([\theta]_2)} \begin{bmatrix} c([\theta]_2) & s([\theta]_1)s([\theta]_2) & c([\theta]_1)s([\theta]_2) \\ 0 & c([\theta]_1)c([\theta]_2) & -s([\theta]_1)c([\theta]_2) \\ 0 & s([\theta]_1) & c([\theta]_1) \end{bmatrix},$$

$$\hat{\omega} = \begin{bmatrix} 0 & [\omega]_3 & -[\omega]_2 \\ -[\omega]_3 & 0 & [\omega]_1 \\ [\omega]_2 & -[\omega]_1 & 0 \end{bmatrix},$$

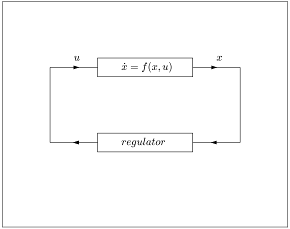
Quadratic cost functional: stabilize the reorientation.

The control error $\|u^k - \bar{u}^N\|_\infty$ versus the number of iterations for different values of $\hat{\omega}$ and N.



Optimal feedback control

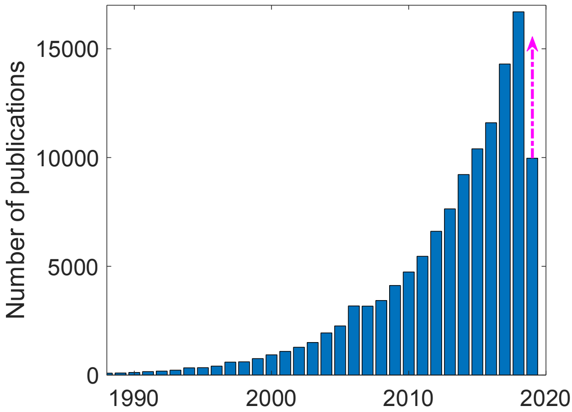
A mapping $(t, x) \mapsto u(t, x)$ such that when applied to the plant, the resulting (x, u) is optimal for every initial state



Under coercivity and isolatedness, there exists an optimal feedback control which is a locally Lipschitz continuous function.

Model Predictive Control (MPC)

MPC: an algorithmic approximation of the optimal feedback control



Thanks to Ilya Kolmanovsky

Model predictive control: Infinite horizon

Finite horizon

The MPC algorithm

1. Discretize the problem.
2. Assume that a control u^N is already determined on $[0, t_k)$;
3. Solve the discrete-time optimal control problem

$$\min \sum_{i=k}^{N-1} h\varphi(x_i, u_i)$$

subject to

$$x_{i+1} = x_i + hg(x_i, u_i), \quad u_i \in U, \quad i = k, \dots, N-1, \quad x_k := x_k^0,$$

obtaining an optimal discrete-time control $(\tilde{u}_k, \dots, \tilde{u}_{N-1})$;
4. Define the constant in time function

$$u^N(t) = \tilde{u}_k \quad \text{for } t \in [t_k, t_{k+1}),$$

and apply it to the continuous-time system on $[t_k, t_{k+1}]$. Measure the value x_{k+1}^0 of the resulting state at t_{k+1} with error ξ_k. Change k to $k+1$ and go to 1.

MPC involves discrete approximations, perturbations, optimization

– discrete approximation: error bounds in terms of the step size h?

– the initial state and the optimization interval change at each step: does the solution depends continuously on the initial state?

– repeated solving of similar optimization problems: does the method chosen behave similarly throughout the iterations? Do we need to always solve the optimization problems exactly?

– MPC-generated control: what is the difference between MPC-generated and an optimal feedback control?

Answers could be found by employing strong regularity.

The MPC-generated control is an approximation of the optimal feedback control

u^N — the MPC-generated control
x^N — the state trajectory of the continuous-time system obtained for u^N
x^f — the state trajectory obtained by applying the optimal feedback u^f
$\hat{u}(t) := u^f(t, x^f(t))$

Theorem.

Under coercivity and isolatedness, there exist a natural N_0 and positive constants ε_0 and c such that for every $N \geq N_0$ and for every measurement error ξ with $\max_{0 \leq i \leq N-1} |\xi_i| \leq \varepsilon_0$, the following estimate holds:

$$\|u^N - \hat{u}\|_{L^1} + \|x^N - x^f\|_{W^{1,1}} \leq c\Big(h + h\sum_{i=1}^{N-1} |\xi_i|\Big).$$

Closing remarks

– Finding (generalized) derivatives of solutions remains a main goal in sensitivity analysis, but these are not always available. Instead, one may directly employ quantitative (Lipschitz) properties. Strong regularity could be particularly helpful for determining Lipschitz constants;

— strong regularity can be used to not only evaluate the effect of parameter perturbations but also to find error estimates for various approximations of the problem, as well as obtaining (uniform) convergence of algorithms;

– strong regularity is just one regularity property of set-valued mappings involved in optimization problems; there are several other regularities such as metric regularity, subregularity, strong subregularity, that are not fully explored yet.

Some references

– AD, R.T. Rockafellar, Implict functions and solution mappings, 2nd edition, Springer 2014.
– A. D. Ioffe, Variational analysis of regular mappings, Springer 2017.
– AD, I. Kolmanovsky, M. Krastanov, M. Nicotra, V. Veliov, Lipschitzian stability in discretized optimal control, SIAM J. Control and Optim., 57 (2019) 468–489.
– AD, I. Kolmanovsky, M. Krastanov, V. Veliov, P. Vuong, Approximating optimal finite-horizon feedback by model predictive control, Systems and Control Letters, 139 (2020).

Thank You!