Characterizations of Input-to-State Stability in Nonlinear Optimization Algorithms

Torbjørn Cunis

University of Stuttgart
Institute of Flight Mechanics and Control

COLLEGE OF ENGINEERING
AEROSPACE ENGINEERING
UNIVERSITY OF MICHIGAN

Seminar *Input-to-State Stability and its Applications*, Online
May 21, 2025

Optimization is not a Singular Event

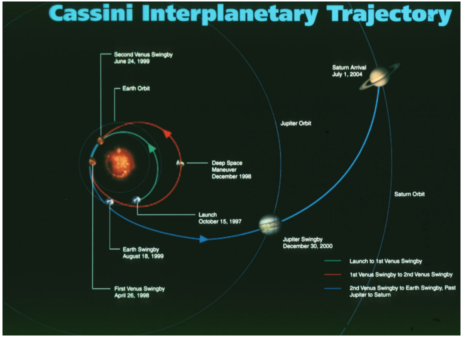


Figure: Interplanetary maneuver. (Source: NASA)

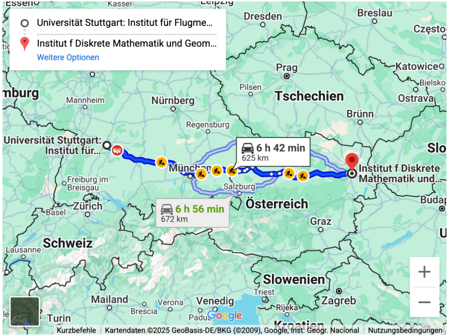
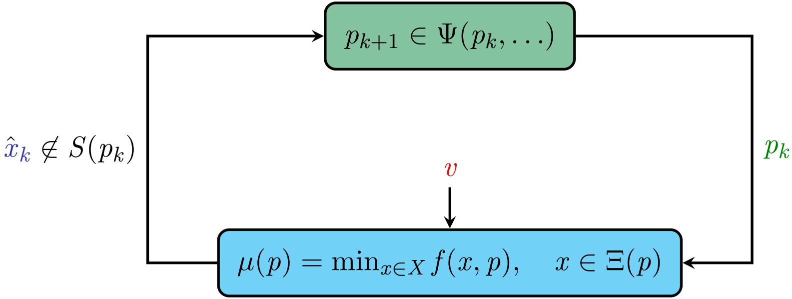


Figure: Planetary maneuver. (Source: Google Maps)

Optimization in (Feedback) Loops

- dynamical system (MPC)
- central unit (distributed OP)
- hyper-parameter optimization
- any other upper-level OP



- optimal-control problem
- convex / unconstrained OP
- neural network training
- any other lower-level OP

Examples

- Model predictive control

$$
\begin{aligned}
u_k &\in \arg\min_u \operatorname{OCP}(u, x_k) \\
x_{k+1} &= f(x_k, u_k)
\end{aligned}
$$

- Hyper-parameter training

$$
\begin{aligned}
\theta_k &\in \arg\min_\theta \operatorname{LSQ}(\theta, p_k) \\
p_{k+1} &= p_k - \partial \operatorname{CVA}(p_k, \theta_k)
\end{aligned}
$$

- Augmented Lagrangian Method

$$
\begin{aligned}
x_k &\in \arg\min_x L_\varrho(x, y_k) \\
y_{k+1} &= y_k - \varrho f(x_k)
\end{aligned}
$$

Input-to-state Stability

Question:

- How does a disturbance affect the interconnected dynamics?

Definition

A dynamic system (e.g., $x_{k+1} = \phi(x_k, v)$) is *input-to-state stable* (ISS) iff

0-GAS: the equilibrium $\bar{x}$ is globally asymptotically stable for $v \equiv 0$;

AG: $\exists \gamma \in \mathcal{K}$, $\forall x_0$, $\forall v \in \ell_\infty$, $\limsup_{N\to\infty} \|x_N\| \leq \gamma(\sup_{k\geq 0} \|v_k\|)$.

Interconnections of ISS Systems

- Let ψ_1, ψ_2 be ISS with gains $\gamma_1, \gamma_2 \in \mathcal{K}$

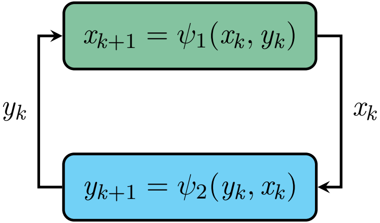


Figure: ISS if $\gamma_1 \circ \gamma_2 \prec \text{id}$

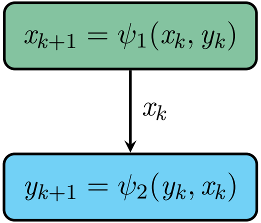


Figure: always ISS

ISS of Optimization Loops

If the optimization algorithm is (locally) ISS, we can prove...

- that an *MPC feedback asymptotically stabilizes* with a finite number of iterations [LMNK20]
- that a *gradient-based bilevel scheme converges* with inexact lower-level solutions and gradients [CK23], even without differentiability in the lower level [CK24a]

ISS of Optimization Algorithms

Algorithms that are (locally) ISS to disturbances:

[HKC13], [CDS20] Newton-like methods for equation systems

[LMNK20] a class of q-linearly convergent algorithms for optimal control

[Son22] gradient descent with Polyak-Łojasiewicz (PL) condition

[CK23] proximal gradient descent with strong convexity or PL

[dOSS23] Newton's method for gradient systems

[CK24b] Josephy-Newton methods for generalized equations including SQP and augmented Lagrangian methods

Brief Introduction: Variational Analysis

Consider the problem

$$\min_x \varphi(x) \quad \text{subject to } x \in C \tag{P}$$

with $\varphi : X \to \mathbb{R}$ continuously (Fréchet) differentiable and $C \subset X$ closed and convex.

Let $\bar{x}$ be a *local minimum*

- that is, for a neighbourhood U of $\bar{x}$,

$$\forall x \in C \cap U, \quad \varphi(x) - \varphi(\bar{x}) \geq 0$$

- then [Don21]

$$\forall x \in C, \quad \nabla\varphi(\bar{x})(x - \bar{x}) \geq 0 \tag{VI}$$

Brief Introduction: Generalized Equations

The necessary conditions (VI) are equivalent to

$$\nabla\varphi(\bar{x}) + N(\bar{x}, C) \ni 0 \tag{GE}$$

where $N(\cdot, C) : x \mapsto \mathcal{N}_x \subset X^*$ is the normal cone mapping.

- $\big[\nabla\varphi + N(\cdot, C)\big] : X \rightrightarrows X^*$ is a *set-valued mapping* (SVM)
- similarly, Karush–Kuhn–Tucker (KKT) conditions with $y \in Y^*$ can be written as $F : X \times Y^* \rightrightarrows X^* \times Y$
- nonlinear optimization algorithms often solve (GE) in lieu of (P)

Brief Introduction: Newton Methods

1. Primal problem and its necessary conditions

$$\min_x \varphi(x) \quad \text{s.t. } x \in C \qquad\qquad \nabla\varphi(x) + N(x, C) \ni 0$$

2. Approximate at $x_k \in X$

$$\min_x \tfrac{1}{2}\nabla^2\varphi(x_k)(x - x_k, x - x_k) + \nabla\varphi(x_k)(x - x_k) \quad \text{s.t. } x \in C \qquad\qquad \nabla\varphi(x_k) + \nabla^2\varphi(x_k)(x - x_k) + N(x, C) \ni 0$$

3. Solve for next iterate

$$x_{k+1} \in \left[\nabla^2\varphi(x_k) + N(\cdot, C)\right]^{-1}\left(\nabla^2\varphi(x_k)x_k - \nabla\varphi(x_k)\right)$$

and repeat.

Brief Introduction: Regularity of SVMs

Newton methods for (perturbed) optimization:

- solve $F(x, v) \ni 0$ through iteration $x_{k+1} \in \Phi(x_k, v_k)$

Idea of (strong or metric) regularity: F and Φ behave 'nicely' around $\bar{x}$

Remark

Notions of *regularity* include (imply)

1. surjectivity or openness
2. injectivity
3. nonsingular linear operator

and these notions are stable under perturbation v.

Outline

Optimization algorithms under strong regularity

Strong regularity in nonlinear optimization

Systems-theoretical characterization of strong regularity

Problem Setup

Consider the nonlinear optimization problem

$$\min_x \varphi(x) \quad \text{s.t.} \;\; g(x) = 0 \text{ and } x \in \Omega \tag{1}$$

with $\varphi : X \to \mathbb{R}$, $g : X \to Y$, and $\Omega \subseteq X$

Assume that

1. X is Asplund and $\bar{x} \in X$ is local optimal solution
2. φ and g are continuously Fréchet differentiable (a fortiori strictly differentiable) around $\bar{x}$
3. $\Omega \subseteq X$ is nonempty, closed, and convex
4. either Ω or $\{0\}$ is SNC at $\bar{x}$ or 0, respectively

Problem Setup

Consider the nonlinear optimization problem

$$\min_x \varphi(x) \quad \text{s.t.} \;\; g(x) = 0 \text{ and } x \in \Omega \tag{1}$$

with $\varphi : X \to \mathbb{R}$, $g : X \to Y$, and $\Omega \subseteq X$

- The *Karush–Kuhn–Tucker* (KKT) necessary conditions for (1) are

$$F(x, y) = \begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix} + \begin{bmatrix} N(x, \Omega) \\ \{0\} \end{bmatrix} \ni 0 \tag{2}$$

  with duals $y \in Y^*$ and normal cone $N(\cdot, \Omega) : X \rightrightarrows X^*$

Example

A *discrete-time optimal control problem* is given as

$$\min_{x = \binom{\xi_1,\ldots,\xi_N}{\upsilon_0,\ldots,\upsilon_{N-1}}} \underbrace{\sum_{k=0}^{N} \ell(\xi_k, \upsilon_k)}_{=\varphi(x)} \quad \text{s.t.} \quad \underbrace{\begin{pmatrix} \xi_1 - \psi(\xi_0, \upsilon_0) \\ \vdots \\ \xi_N - \psi(\xi_{N-1}, \upsilon_{N-1}) \end{pmatrix}}_{=g(x)} = 0 \text{ and } x \in \underbrace{\mathcal{X} \times \mathcal{U}}_{=\Omega}$$

where $\mathcal{X}$ and $\mathcal{U}$ are state and input constraint sets (polygonal or hyperboxes)

This problem is a *nonlinear program* (NLP):

- X and Y are finite-dimensional
- φ and g are (usually) twice differentiable
- Ω is given by linear constraints

Generalized Newton Methods

Solve the KKT generalized equation

$$\underbrace{\begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix}}_{= f(z)} + \underbrace{\begin{bmatrix} N(x,\Omega) \\ \{0\} \end{bmatrix}}_{= N(z)} \ni 0$$

via the iteration

$$f(z_k) + H(z_k)(z_{k+1} - z_k) + N(z_{k+1}) \ni 0 \tag{3}$$

for a suitable operator $H(\cdot)$ and $z = (x, y)$

Generalized Newton Methods

Solve the KKT generalized equation

$$\underbrace{\begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix}}_{= f(z)} + \underbrace{\begin{bmatrix} N(x, \Omega) \\ \{0\} \end{bmatrix}}_{= N(z)} \ni 0$$

via the iteration

$$f(z_k) + H(z_k)(z_{k+1} - z_k) + N(z_{k+1}) \ni 0 \tag{3}$$

Remark (Sequential quadratic programming)

$$H(z) = \begin{pmatrix} \nabla^2(\varphi(x) + \langle g(x), y\rangle) & \nabla g(x)^* \\ \nabla g(x) & 0 \end{pmatrix}$$

Generalized Newton Methods

Solve the KKT generalized equation

$$\underbrace{\begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix}}_{= f(z)} + \underbrace{\begin{bmatrix} N(x,\Omega) \\ \{0\} \end{bmatrix}}_{= N(z)} \ni 0$$

via the iteration

$$f(z_k) + H(z_k)(z_{k+1} - z_k) + N(z_{k+1}) \ni 0 \tag{3}$$

Remark (Sequential linear programming)

$$H(z) = \begin{pmatrix} 0 & \nabla g(x)^* \\ \nabla g(x) & 0 \end{pmatrix}$$

Generalized Newton Methods

Solve the KKT generalized equation

$$\underbrace{\begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix}}_{= f(z)} + \underbrace{\begin{bmatrix} N(x,\Omega) \\ \{0\} \end{bmatrix}}_{= N(z)} \ni 0$$

via the iteration

$$f(z_k) + H(z_k)(z_{k+1} - z_k) + N(z_{k+1}) \ni 0 \tag{3}$$

Remark (Projected gradient)

$$H(z) = \begin{pmatrix} \alpha^{-1}\mathbb{I} & 0 \\ 0 & \alpha^{-1}\mathbb{I} \end{pmatrix}$$

Perturbed Newton Methods

Solve the *perturbed* generalized equation

$$f(z, v) + N(z) \ni 0$$

via the perturbed generalized Newton iteration

$$z_{k+1} \in \Phi_H(z_k, v_k) \iff_{\text{def}} f(z_k, v_k) + H(z, v_k)(z_{k+1} - z_k) + N(z_{k+1}) \ni 0$$

for a suitable operator $H(\cdot)$

Strong Regularity

Definition

F is *strongly regular* at $\bar{z}$ for $\bar{v} \in F(\bar{z})$ iff, with neighbourhoods U of $\bar{z}$ and V of $\bar{v}$,

$$\forall v \in V, \quad F^{-1}(v) \cap U = \{s(v)\}$$

and $s(\cdot)$ is Lipschitz continuous around $\bar{v}$.

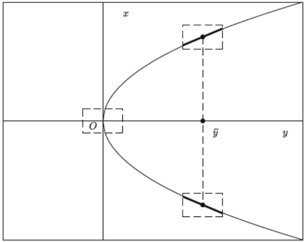


Figure: The inverse of $z \mapsto z^2$.

Equivalent: [Don21]

1. F is strongly regular at $\bar{z}$ for $\bar{v}$,
2. F^{-1} has Lipschitz continuous, single-valued localization at $\bar{v}$ for $\bar{z}$
3. F is linearly open (a fortiori surjective) and locally injective at $\bar{z}$ for $\bar{v}$

ISS of Newton Methods

Let $f_H(z', z, v) = f(z, v) + H(z, v)(z' - z)$;

Theorem ([CK24b])

Suppose that

1. $\bar{z}$ *is a solution of* $f(\cdot, 0) + N \ni 0$
2. f_H *is uniformly Lipschitz continuous (constants* γ_z *and* γ_v*) at* $(\bar{z}, \bar{z}, 0)$
3. $f_H(\cdot, \bar{z}, 0) + N$ *is strongly regular (constant* κ*) at* $\bar{z}$ *for* 0

and $\kappa\gamma_z < 1$*; then the iteration* $z_{k+1} \in \Phi(z_k, v)$
is locally unique and locally input-to-state stable.

Proof sketch: The update has a locally unique solution $s(\cdot)$ with

$$\|z_{k+1} - \bar{z}\| = \|s(z_k, v_k) - s(\bar{z}, 0)\| \leq \kappa\gamma_z\|z_k - \bar{z}\| + \kappa\gamma_v\|v_k\|$$

Generalized Implicit Function Theorem

Let $f_H(z', p) = f(p) + H(p)(z' - p_1)$ with $p = (z, v)$;

Proposition ([Don21])

Suppose that

1. $\bar{z}$ *is a solution of* $f_H(\cdot, \bar{p}) + N \ni 0$
2. f_H *is uniformly Lipschitz continuous (constant* γ_p*) at* $(\bar{z}, \bar{p})$
3. $f_H(\cdot, \bar{p}) + N$ *is strongly regular (constant* κ*) at* $\bar{z}$ *for* 0

then

$$S : p \mapsto \{z \in X \times Y^* \mid f_H(z, p) + N(z) \ni 0\}$$

has a Lipschitz continuous (constant $\kappa\gamma_p$*) and single-valued localization* $s(\cdot)$ *at* $\bar{p}$ *for* $\bar{z}$*.*

Strong Regularity in Nonlinear Optimization

The mapping

$$f_H(\cdot, \bar{p}) + N : (x, y) \mapsto f(\bar{p}) + \begin{bmatrix} H_{xx} & H_{yx}^* \\ H_{yx} & 0 \end{bmatrix} \begin{pmatrix} x - \bar{x} \\ y - \bar{y} \end{pmatrix} + N(x, \Omega)$$

with $H_{xx} \succeq 0$ is strongly regular at $\bar{z}$ for 0 *if and only if*

$$\min_x H_{xx}(x - \bar{x}, x - \bar{x}) + [\nabla \varphi(\bar{x}) - \delta_x](x - \bar{x})$$
$$\text{s.t.} \;\; g(\bar{x}) + H_{yx}(x - \bar{x}) = \delta_y \text{ and } x \in \Omega$$

has a unique primal-dual solution (x_δ, y_δ) for $\delta = (\delta_x, \delta_y)$ close to 0 with

$$\|(x_{\delta 1}, y_{\delta 1}) - (x_{\delta 2}, y_{\delta 2})\| \leq \kappa \|\delta_1 - \delta_2\|$$

Nonlinear Programs revisited

Consider the nonlinear program

$$\min_x \varphi(x) \quad \text{s.t. } g(x) = 0 \text{ and } x \in \Omega = \mathbb{R}^n_{\geq 0} \tag{4}$$

with $\varphi : X \to \mathbb{R}$, $g : X \to Y$, and $\Omega \subseteq X$

Assume that

1. X and Y are finite-dimensional; and $\bar{x} \in X$ is local optimal solution
2. φ and g are *twice* continuously Fréchet differentiable
3. $\Omega = \mathbb{R}^n_{\geq 0}$ is the nonnegative orthant

Note: $N(x, \mathbb{R}^n_{\geq 0}) \subseteq \mathbb{R}^n_{\leq 0}$

Constraint Qualifications

MFCQ

If the constraint qualification

$$\{\nabla g(\bar{x})^* y \mid y \in Y^*\} \cap \left[-N(\bar{x}, \mathbb{R}^n_{\geq 0})\right] = \{0\}$$

holds and $\nabla g(\bar{x})$ is surjective, then $F(\bar{x}, \bar{y}) = 0$ for *some* $\bar{y} \in Y^*$.

LICQ

If the constraint qualification

$$\{\nabla g(\bar{x})^* y \mid y \in Y^*\} \cap \left[\operatorname{span} N(\bar{x}, \mathbb{R}^n_{\geq 0})\right] = \{0\}$$

holds and $\nabla g(\bar{x})$ is surjective, then $F(\bar{x}, \bar{y}) = 0$ for a unique $\bar{y} \in Y^*$.

Strong Stability in Nonlinear Programs

$\bar{x} \in X$ is a *stationary solution* if and only if $F(\bar{x}, \bar{y}) = 0$ for some $\bar{y} \in Y^*$.

Definition ([Koj80])

A stationary solution $\bar{x}$ is *strongly stable* if and only if there exists a neighbourhood U of $\bar{x}$ and $d > 0$ such that

$$\min_x \varphi(x) + \langle \Delta x + \delta_\varphi, x \rangle \quad \text{s.t.} \;\; g(x) = \delta_y \text{ and } x + \delta_x \in \mathbb{R}^n_{\geq 0}$$

has a unique stationary solution $s(\cdot) \in U$ for $\|(\Delta, \delta_\varphi, \delta_y, \delta_x)\| \leq d$ which is continuous at 0.

Strong Stability & Strong Regularity

If F is strongly regular at $(\bar{x}, \bar{y})$ for 0, then

- $\bar{x}$ is a strongly stable stationary solution
- MFCQ holds at $\bar{x}$ with unique dual $\bar{y}$

An *optimal* solution $\bar{x}$ is a strongly stable stationary solution if and only if

- MFCQ holds
- the strong second-order sufficient condition is satisfied (a fortiori, $\bar{x}$ is a strict local minimum)

If $\bar{x}$ is a strongly stable stationary solution and LICQ holds, then

- F is strongly regular at $(\bar{x}, \bar{y})$ for 0

Towards a Systems-theoretical Characterization

Consider the *canonically perturbed* optimization problem

$$\min_x \varphi(x) - \langle v_x, x \rangle \quad \text{s.t.} \; g(x) = v_y \text{ and } x \in \Omega \tag{5}$$

with $\varphi : X \to \mathbb{R}$, $g : X \to Y$, and $\Omega \subseteq X$ for $(v_x, v_y) \in X^* \times Y$

- The KKT conditions for (5) become

$$F(x, y) = \underbrace{\begin{pmatrix} \nabla\varphi(x) + \nabla g(x)^* y \\ g(x) \end{pmatrix}}_{= f(z)} + \underbrace{\begin{bmatrix} N(x, \Omega) \\ \{0\} \end{bmatrix}}_{= N(z)} \ni \underbrace{\begin{pmatrix} v_x \\ v_y \end{pmatrix}}_{= v} \tag{6}$$

  with duals $y \in y$ and normal cone $N(\cdot, \Omega) : X \rightrightarrows X^*$

Necessary Conditions for ISS

Suppose that the generalized Newton's iteration

$$z_{k+1} \in \Phi_H(z_k, v_k) \iff f(z_k) + H(z_k)(z_{k+1} - z_k) - v_k + N(z_k) \ni 0$$

is locally input-to-state stable around $\bar{z} \in F^{-1}(0)$, that is,

$$\|z_N - \bar{z}\| \leq \alpha^N \|z_0 - \bar{z}\| + \gamma \sup_{k \geq 0} \|v_k\|$$

for all $z_0 \in U$, $z_{k+1} \in \Phi_H(z_k, v_k) \cap U$, $v_k \in V$, and $N \geq k \geq 0$, where $\alpha \in (0, 1)$ and $\gamma \geq 0$

- Any fixpoint $z_v \in \Phi_H(z_v, v) \cap U$ for $v \in V$ satisfies

$$\|z_v - \bar{z}\| \leq \gamma \|v\|$$

Sufficient Conditions for Strong Subregularity

Suppose that the generalized Newton's iteration

$$z_{k+1} \in \Phi_H(z_k, v_k) \iff f(z_k) + H(z_k)(z_{k+1} - z_k) - v_k + N(z_k) \ni 0$$

is locally input-to-state stable around $\bar{z} \in F^{-1}(0)$, that is,

$$\|z_N - \bar{z}\| \leq \alpha^N \|z_0 - \bar{z}\| + \gamma \sup_{k \geq 0} \|v_k\|$$

- Any solution $z_v \in (f + N)^{-1}(v) \cap U$ for $v \in V$ satisfies

$$\|z_v - \bar{z}\| \leq \gamma \|v\|$$

- Hence, $f + N$ is *strongly subregular* at $\bar{z}$ for 0

Sufficient Conditions for Strong Regularity

Theorem (work in progress)

If the generalized Newton's iteration

$$z_{k+1} \in \Phi_H(z_k, v_k) \iff f(z_k) + H(z_k)(z_{k+1} - z_k) - v_k + N(z_k) \ni 0$$

1. *has a fix point* $z_v \in \Phi_H(z_v, v) \cap U$ *for all* $v \in V$ *and*
2. *is locally* incrementally *ISS around* $\bar{z} \in F^{-1}(0)$, *that is,*

$$\|z'_N - z_N\| \leq \alpha^N \|z'_0 - z_0\| + \gamma \sup_{k \geq 0} \|v'_k - v_k\|$$

for all $z_0^{(\prime)} \in U$, $z_{k+1}^{(\prime)} \in \Phi_H(z_k^{(\prime)}, v_k^{(\prime)}) \cap U$, $v_k^{(\prime)} \in V$, $N \geq k \geq 0$, *where* $\alpha \in (0, 1)$ *and* $\gamma \geq 0$,

Then $f + N$ *is strongly regular at* $\bar{z}$ *for 0.*

Concluding Remarks

Regularity of the KKT conditions

- impacts sensitivity and stability in nonlinear optimization
- relates to stable stationary solutions and second-order sufficiency conditions in nonlinear programs (NLP) *and beyond*
- applies to nonlinear optimization problems other than NLPs e.g., nonconvex semidefinite or sum-of-squares programs

Beyond Strong Regularity

- Strong regularity implies local incremental ISS and uniqueness of primal-dual solutions under perturbations

Alternatives:

1. Strong subregularity (implies ISS)
2. Metric regularity (implies existence)

Remark

Strong regularity and metric regularity are equivalent for NLPs.

Acknowledgments





Thank you!

`tcunis@ifr.uni-stuttgart.de`
`tcunis@umich.edu`

References I

Giuseppe G. Colabufo, Peter M. Dower, and Iman Shames, *Newton's method: Sufficient conditions for practical and input-to-state stability* , IFAC-PapersOnLine **53** (2020), no. 2, 6334–6339.

Torbjørn Cunis and Ilya Kolmanovsky, *Input-to-State Stability of a Bilevel Proximal Gradient Descent Algorithm* , IFAC-PapersOnLine **56** (2023), no. 2, 7474–7479.

______, *Inexactness in Bilevel Nonlinear Optimization: A Gradient-free Newton's Method Approach* , Symposium on Systems Theory in Data and Optimization, 5 2024.

______, *Input-to-State Stability of Newton Methods for Generalized Equations in Nonlinear Optimization* , 2024 IEEE Conference on Decision and Control (Milano), 3 2024.

References II

- Asen L. Dontchev, *Lectures on Variational Analysis* , Applied Mathematical Sciences, no. 205, Springer, Cham, 2021.
- Arthur Castello B. de Oliveira, Milad Siami, and Eduardo D. Sontag, *Dynamics and Perturbations of Overparameterized Linear Neural Networks* , Proceedings of the IEEE Conference on Decision and Control, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 7356–7361.
- Ammar Hasan, Eric C. Kerrigan, and George A. Constantinides, *Control-theoretic forward error analysis of iterative numerical algorithms* , IEEE Transactions on Automatic Control **58** (2013), no. 6, 1524–1529.
- Masakazu Kojima, *Strongly Stable Stationary Solutions in Nonlinear Programs* , Analysis and Computation of Fixed Points (Stephen M. Robinson, ed.), Academic Press, New York, NY, 1980, pp. 93–138.

References III

Dominic Liao-McPherson, Marco M. Nicotra, and Ilya Kolmanovsky, *Time-distributed optimization for real-time model predictive control: Stability, robustness, and constraint satisfaction* , Automatica **117** (2020), no. October, 108973.

Eduardo D. Sontag, *Remarks on input to state stability of perturbed gradient flows, motivated by model-free feedback control learning* , Systems and Control Letters **161** (2022), 105138.