A Survey of Generalized Gauss-Newton and Sequential Convex Programming Methods

Moritz Diehl

Systems Control and Optimization Laboratory
Department of Microsystems Engineering and Department of Mathematics
University of Freiburg, Germany

based on joint work with
Florian Messerer (Freiburg)
and Joris Gillis (Leuven)

19th French-German-Swiss Conference on Optimization, Nice, September 18, 2019

A Tutorial of Generalized Gauss-Newton and Sequential Convex Programming Methods

Moritz Diehl

Systems Control and Optimization Laboratory
Department of Microsystems Engineering and Department of Mathematics
University of Freiburg, Germany

based on joint work with
Florian Messerer (Freiburg)
and Joris Gillis (Leuven)

19th French-German-Swiss Conference on Optimization, Nice,
September 18, 2019

Nonlinear optimization with convex substructure

$$\begin{aligned} \underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & \phi_0(F_0(w)) \\ \text{subject to} \quad & F_i(w) \in \Omega_i \quad i = 1, \dots, m, \\ & G(w) = 0 \end{aligned}$$

Assumptions:

- twice continuously differentiable functions $G : \mathbb{R}^{n_w} \to \mathbb{R}^{n_g}$ and $F_i : \mathbb{R}^{n_w} \to \mathbb{R}^{n_{F_i}}$ for $i = 0, 1, \dots, m$.
- outer function $\phi_0 : \mathbb{R}^{n_{F_0}} \to \mathbb{R}$ convex.
- sets $\Omega_i \subset \mathbb{R}^{n_{F_i}}$ convex for $i = 1, \dots, m$,
  (possibly $z \in \Omega_i \Leftrightarrow \phi_i(z) \leq 0$ with smooth convex ϕ_i)

Nonlinear optimization with convex substructure

$$\begin{aligned} \underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & \phi_0(F_0(w)) \\ \text{subject to} \quad & F_i(w) \in \Omega_i \quad i = 1, \ldots, m, \\ & G(w) = 0 \end{aligned}$$

Assumptions:

- twice continuously differentiable functions $G : \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_g}$ and $F_i : \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_{F_i}}$ for $i = 0, 1, \ldots, m$.
- outer function $\phi_0 : \mathbb{R}^{n_{F_0}} \rightarrow \mathbb{R}$ convex.
- sets $\Omega_i \subset \mathbb{R}^{n_{F_i}}$ convex for $i = 1, \ldots, m$,
  (possibly $z \in \Omega_i \Leftrightarrow \phi_i(z) \leq 0$ with smooth convex ϕ_i)

Idea:
exploit convex substructure via *iterative convex approximations*.

Why is this class of problems and algorithms interesting?

- some nonlinear programming (NLP) problems have nonsmooth convex constraints which cannot be treated by standard NLP solvers
- there exist many reliable and efficient convex optimization solvers

Some application areas:

- nonlinear matrix inequalities for reduced order controller design [Fares, Noll, Apkarian 2002; Tran-Dinh et al. 2012]
- ellipsoidal terminal regions in nonlinear model predictive control [Chen and Allgöwer 1998; Verschueren 2016]
- robustified inequalities in nonlinear optimization [Nagy and Braatz 2003; D., Bock, Kostina 2006]
- tube-following optimal control problems [Van Duijkeren, 2019]
- non-smooth composite minimization [Lewis and Wright 2016]
- deep neural network training with convex loss functions [Schraudolph 2002; Martens 2016]

First the bad news

First the bad news

- Iterative convex approximation methods such as sequential convex programming (SCP) have **only linear convergence** in general.

First the bad news

- Iterative convex approximation methods such as sequential convex programming (SCP) have **only linear convergence** in general.
- The rate of convergence cannot be improved to superlinear by any bounded semi-definite Hessian approximation [D., Jarre, Vogelbusch 2006]

Simple TR example problem with dominant nonconvexities in objective:

$$\begin{aligned} \underset{w \in \mathbb{R}^2}{\text{minimize}} \quad & -w_1^2 - (1-w_2)^2 \\ \text{subject to} \quad & \|w\|_2 \le 1 \end{aligned}$$

First the bad news

- Iterative convex approximation methods such as sequential convex programming (SCP) have **only linear convergence** in general.
- The rate of convergence cannot be improved to superlinear by any bounded semi-definite Hessian approximation [D., Jarre, Vogelbusch 2006]

Simple TR example problem with dominant nonconvexities in objective:

$$\begin{aligned} \underset{w \in \mathbb{R}^2}{\text{minimize}} \quad & -w_1^2 - (1-w_2)^2 \\ \text{subject to} \quad & \|w\|_2 \leq 1 \end{aligned}$$

But: many real-world problems have dominant convexities and SCP often shows fast linear convergence in practice.

First the bad news

- Iterative convex approximation methods such as sequential convex programming (SCP) have **only linear convergence** in general.
- The rate of convergence cannot be improved to superlinear by any bounded semi-definite Hessian approximation [D., Jarre, Vogelbusch 2006]

Simple TR example problem with dominant nonconvexities in objective:

$$\begin{aligned} \underset{w \in \mathbb{R}^2}{\text{minimize}} \quad & -w_1^2 - (1-w_2)^2 \\ \text{subject to} \quad & \|w\|_2 \leq 1 \end{aligned}$$

But: many real-world problems have dominant convexities and SCP often shows fast linear convergence in practice. How fast?

Overview

- Smooth unconstrained problems
  - Sequential Convex Programming (SCP)
  - Generalized Gauss-Newton (GGN)
  - Local convergence analysis
  - Desirable divergence
  - Illustrative example in CasADi
- Constrained problems
  - Sequential Convex Programming (SCP)
  - Sequential Linear Programming (SLP)
  - Constrained Generalized Gauss-Newton (CGGN)
  - Sequential Convex Quadratic Programming (SCQP)
- Recent Progress in Dynamic Optimization and Applications

Simplest case: smooth unconstrained problems

Unconstrained minimization of "convex over nonlinear" function:

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \quad \underbrace{\phi(F(w))}_{=: f(w)}$$

Assumptions:
- Inner function $F : \mathbb{R}^n \to \mathbb{R}^N$ of class C^3
- Outer function $\phi : \mathbb{R}^N \to \mathbb{R}$ of class C^3 and convex

Remark:
$F, \phi, f \in C^3$ avoids technical details that would obfuscate main results.

Notation and Preliminaries

Matrix inequality $A \succ B$ for $A, B \in \mathbb{S}^n$ (symmetric matrices)

Gradient $\nabla f(w) \in \mathbb{R}^n$ and Hessian $\nabla^2 f(w) \in \mathbb{S}^n$ for $f : \mathbb{R}^n \to \mathbb{R}$

Jacobian $J(w) := \frac{\partial F}{\partial w}(w) \in \mathbb{R}^{N \times n}$ for $F : \mathbb{R}^n \to \mathbb{R}^N$

Linearization (first order Taylor series) at $\bar{w} \in \mathbb{R}^n$:

$$F_{\text{lin}}(w; \bar{w}) := F(\bar{w}) \; + \; J(\bar{w}) \; (w - \bar{w})$$

Big $O(\cdot)$ notation: e.g. $F_{\text{lin}}(w; \bar{w}) = F(w) + O(\|w - \bar{w}\|^2)$

Note: gradient of $f(w) = \phi(F(w))$ is $\nabla f(w) = J(w)^\top \nabla \phi(F(w))$

Method 1: Sequential Convex Programming (SCP)

Starting at $w_0 \in \mathbb{R}^n$ we generate sequence $\ldots, w_k, w_{k+1}, \ldots$

by obtaining w_{k+1} as solution of convex subproblem at $\bar{w} = w_k$:

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \quad \underbrace{\phi\left(F_{\text{lin}}(w; \bar{w})\right)}_{=: f_{\text{SCP}}(w;\bar{w})} \tag{1}$$

(requires possibly expensive calls of a convex optimization solver)

Observation: $f_{\text{SCP}}(w; \bar{w}) = f(w) + O(\|w - \bar{w}\|^2)$

Corollary: $\nabla f(\bar{w}) = 0 \quad \Leftrightarrow \quad \bar{w}$ fixed point of SCP

(if subproblems (1) have unique solutions)

Tutorial Example

Regard $F(w) = \begin{bmatrix} \sin(w) \\ \exp(w) - 2 \end{bmatrix}$ and $\phi(z) = z_1^4 + z_2^2$

Linearization: $F_{\text{lin}}(w; \bar{w}) = \begin{bmatrix} \sin(\bar{w}) + \cos(\bar{w})(w - \bar{w}) \\ \exp(\bar{w}) - 2 + \exp(\bar{w})(w - \bar{w}) \end{bmatrix}$

SCP objective:

$$\begin{aligned} f_{\text{SCP}}(w; \bar{w}) \quad = \quad & (\sin(\bar{w}) + \cos(\bar{w})(w - \bar{w}))^4 \quad + \\ & (\exp(\bar{w}) - 2 + \exp(\bar{w})(w - \bar{w}))^2 \end{aligned}$$

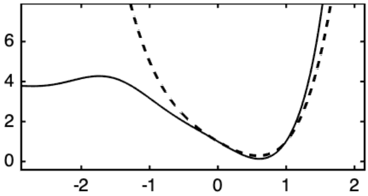


Fig.: $f(w) = \phi(F(w))$ (solid) and $f_{\text{SCP}}(w; \bar{w})$ (dashed) for $\bar{w} = 0$

SCP for Least Squares Problems = Gauss-Newton

With quadratic $\phi(z) = \frac{1}{2}\|z\|_2^2 = \frac{1}{2}z^\top z$, SCP subproblems become

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|F(w_k) + J(w_k)(w - w_k)\|_2^2 \tag{2}$$

If $\text{rank}(J) = n$ this is uniquely solvable, giving

$$w_{k+1} = w_k - \Big(\underbrace{J(w_k)^\top J(w_k)}_{=:B_{\text{GN}}(w_k)}\Big)^{-1} \underbrace{J(w_k)^\top F(w_k)}_{=\nabla f(w_k)}$$

SCP applied to LS = Newton method with "Gauss-Newton Hessian"

$$B_{\text{GN}}(w) \approx \nabla^2 f(w)$$

Method 2: Generalized Gauss-Newton cf. [Schraudolph 2002]

For general convex $\phi(\cdot)$ we have for $f(w) = \phi(F(w))$

$$\nabla^2 f(w) = \underbrace{J(w)^\top \, \nabla^2 \phi(F(w)) \, J(w)}_{=:B_{\mathrm{GGN}}(w)} + \underbrace{\sum_{j=1}^{N} \nabla^2 F_j(w) \, \nabla_{z_j} \phi(F(w))}_{=:E_{\mathrm{GGN}}(w)}$$

"GGN Hessian" "Error matrix"

Method 2: Generalized Gauss-Newton cf. [Schraudolph 2002]

For general convex $\phi(\cdot)$ we have for $f(w) = \phi(F(w))$

$$\nabla^2 f(w) = \underbrace{J(w)^\top \, \nabla^2\phi(F(w)) \, J(w)}_{=: B_{\mathrm{GGN}}(w)} + \underbrace{\textstyle\sum_{j=1}^N \nabla^2 F_j(w) \, \nabla_{z_j}\phi(F(w))}_{=: E_{\mathrm{GGN}}(w)}$$

"GGN Hessian" "Error matrix"

Generalized Gauss-Newton (GGN) method iterates according to

$$w_{k+1} = w_k - B_{\mathrm{GGN}}(w)^{-1}\nabla f(w_k)$$

Method 2: Generalized Gauss-Newton cf. [Schraudolph 2002]

For general convex $\phi(\cdot)$ we have for $f(w) = \phi(F(w))$

$$\nabla^2 f(w) = \underbrace{J(w)^\top \nabla^2\phi(F(w))\, J(w)}_{=:B_{\mathrm{GGN}}(w)} + \underbrace{\textstyle\sum_{j=1}^N \nabla^2 F_j(w)\, \nabla_{z_j}\phi(F(w))}_{=:E_{\mathrm{GGN}}(w)}$$

"GGN Hessian" "Error matrix"

Generalized Gauss-Newton (GGN) method iterates according to

$$w_{k+1} = w_k - B_{\mathrm{GGN}}(w)^{-1}\nabla f(w_k)$$

Note: GGN solves convex quadratic subproblems

$$\min_{w\in\mathbb{R}^n} \underbrace{f(w_k)+\nabla f(w_k)^\top(w-w_k)+\frac{1}{2}(w-w_k)^\top B_{\mathrm{GGN}}(w_k)(w-w_k)}_{=:f_{\mathrm{GGN}}(w;w_k)}$$

Tutorial Example: GGN and SCP

Regard again $F(w) = \begin{bmatrix} \sin(w) \\ \exp(w) - 2 \end{bmatrix}$ and $\phi(z) = z_1^4 + z_2^2$

Jacobian:

$$J(\bar{w}) = \begin{bmatrix} \cos(\bar{w}) \\ \exp(\bar{w}) \end{bmatrix}$$

GGN objective at $\bar{w} = 0$:

$$f_{\mathrm{GGN}}(w; \bar{w}) = \left(\exp(\bar{w}) - 2 + \exp(\bar{w})(w - \bar{w})\right)^2$$

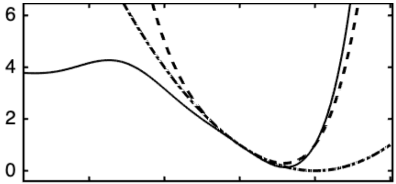


Figure shows f (solid), f_{SCP} (dashed) and f_{GGN} (dash dotted)

Differences and Similarities of SCP and GGN

- both SCP and GGN generalize the Gauss-Newton method (and become equal to it when applied to nonlinear least squares)
- both iteratively solve convex optimization problems
- GGN only solves a positive definite linear system per iteration
- SCP solves a full convex problem per iteration (5-30x slower)
- SCP often less sensitive to initialization
- both are NOT second order methods
- both converge linearly *with the same rate* (details follow)

Local Convergence Analysis for SCP and GGN

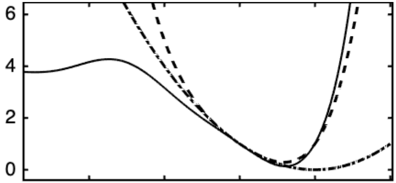
Recall:

$$f_{\text{SCP}}(w;\bar{w}) = \phi\left(F_{\text{lin}}(w;\bar{w})\right)$$
$$f_{\text{GGN}}(w;\bar{w}) = f(\bar{w}) + \nabla f(\bar{w})^\top (w-\bar{w}) + \tfrac{1}{2}(w-\bar{w})^\top B_{\text{GGN}}(\bar{w})(w-\bar{w})$$
$$\nabla_w f_{\text{SCP}}(\bar{w};\bar{w}) = \nabla_w f_{\text{GGN}}(\bar{w};\bar{w}) = \nabla f(\bar{w})$$

Lemma 1 (Equality of SCP and GGN up to Second Order)

$$f_{\text{SCP}}(w;\bar{w}) = f_{\text{GGN}}(w;\bar{w}) + O(\|w-\bar{w}\|^3)$$

Proof: $\nabla_w^2 f_{\text{SCP}}(\bar{w};\bar{w}) = B_{\text{GGN}}(\bar{w}) = \nabla_w^2 f_{\text{GGN}}(\bar{w};\bar{w})$ □



Note: in general $B_{\text{GGN}}(\bar{w}) \neq \nabla^2 f(\bar{w})$

Local Convergence of SCP and GGN

Main Theorem 1 [Diehl and Messerer, CDC 2019]

Regard w^* with $\nabla f(w^*) = 0$ and $B_{\mathrm{GGN}}(w^*) \succ 0$. Then

- w^* is a fixed point for both the SCP and GGN iterations
- both methods are well-defined in a neighborhood of w^*
- their linear contraction rates are equal and given by the LMI

$$\min\{\alpha \in \mathbb{R} \mid -\alpha B_{\mathrm{GGN}}(w^*) \preceq E_{\mathrm{GGN}}(w^*) \preceq \alpha B_{\mathrm{GGN}}(w^*)\}$$

Local Convergence of SCP and GGN

Main Theorem 1 [Diehl and Messerer, CDC 2019]

Regard w^* with $\nabla f(w^*) = 0$ and $B_{\mathrm{GGN}}(w^*) \succ 0$. Then

- w^* is a fixed point for both the SCP and GGN iterations
- both methods are well-defined in a neighborhood of w^*
- their linear contraction rates are equal and given by the LMI

$$\min\{\alpha \in \mathbb{R} \mid -\alpha B_{\mathrm{GGN}}(w^*) \preceq E_{\mathrm{GGN}}(w^*) \preceq \alpha B_{\mathrm{GGN}}(w^*)\}$$

Corollary: Necessary condition for local convergence of both methods is

$$B_{\mathrm{GGN}}(w^*) \succeq \frac{1}{2}\nabla^2 f(w^*) \succeq 0$$

Local Convergence of SCP and GGN

Main Theorem 1 [Diehl and Messerer, CDC 2019]

Regard w^* with $\nabla f(w^*) = 0$ and $B_{\mathrm{GGN}}(w^*) \succ 0$. Then

- w^* is a fixed point for both the SCP and GGN iterations
- both methods are well-defined in a neighborhood of w^*
- their linear contraction rates are equal and given by the LMI

$$\min\{\alpha \in \mathbb{R} \mid -\alpha B_{\mathrm{GGN}}(w^*) \preceq E_{\mathrm{GGN}}(w^*) \preceq \alpha B_{\mathrm{GGN}}(w^*)\}$$

Corollary: Necessary condition for local convergence of both methods is

$$B_{\mathrm{GGN}}(w^*) \succeq \frac{1}{2}\nabla^2 f(w^*) \succeq 0$$

*Proof of corollary: Set $\alpha = 1$ and use $E_{\mathrm{GGN}}(w^*) = \nabla_w^2 f(w^*) - B_{\mathrm{GGN}}(w^*)$

*Proof of Main Theorem

Define solution operators $w^{\mathrm{sol}}_{\mathrm{SCP}}(\bar{w})$ and $w^{\mathrm{sol}}_{\mathrm{GGN}}(\bar{w})$ at linearization point $\bar{w}$ and apply the implicit function theorem to

$$\nabla_w f_i(w_i^{\mathrm{sol}}(\bar{w}); \bar{w}) = 0 \quad \text{for} \quad i = \mathrm{SCP}, \mathrm{GGN}$$

Well-defined for $\bar{w}$ in neighborhood of w^* because $\nabla_w^2 f_{\mathrm{SCP}}(w^*; w^*) = \nabla_w^2 f_{\mathrm{GGN}}(w^*; w^*) = B_{\mathrm{GGN}}(w^*) \succ 0$, and derivatives are given by

$$\frac{\mathrm{d}w_i^{\mathrm{sol}}}{\mathrm{d}\bar{w}}(w^*) = -(\underbrace{\nabla_w \nabla_w f_i(w^*; w^*)}_{=B_{\mathrm{GGN}}(w^*)=:B_*})^{-1} \underbrace{\nabla_{\bar{w}} \nabla_w f_i(w^*; w^*)}_{=E_{\mathrm{GGN}}(w^*)=:E_*}$$

We used that all second derivatives are equal due to Lemma 3 and that

$$\begin{aligned} \nabla_{\bar{w}} \nabla_w f_{\mathrm{GGN}}(w^*; w^*) &= \nabla_{\bar{w}} \left(\nabla f(\bar{w}) + B_{\mathrm{GGN}}(\bar{w})(w - \bar{w})\right)\big|_{w=\bar{w}=w^*} \\ &= \nabla^2 f(w^*) - B_{\mathrm{GGN}}(w^*) = E_{\mathrm{GGN}}(w^*) \end{aligned}$$

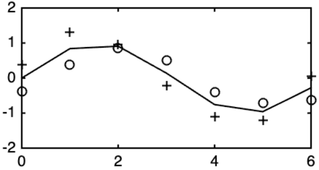
*Proof of Main Theorem (continued)

Spectral radius $\rho(-B_*^{-1}E_*) = \rho(B_*^{-1}E_*)$ equals linear contraction rate of SCP and GGN algorithms. Matrix can be transformed to similar, but symmetric matrix: $B_*^{-1}E_* \sim B_*^{-\frac{1}{2}}\ E_*\ B_*^{-\frac{1}{2}}$ with same spectral radius.

Now, $$\rho(B_*^{-\frac{1}{2}}\ E_*\ B_*^{-\frac{1}{2}}) \le \alpha \quad \Leftrightarrow \quad -\alpha I \preceq B_*^{-\frac{1}{2}}\ E_*\ B_*^{-\frac{1}{2}} \preceq \alpha I$$
$$\Leftrightarrow \quad -\alpha B_* \preceq E_* \preceq \alpha B_* \qquad \square$$

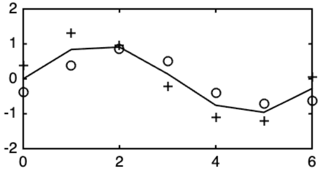
Desirable divergence and mirror problem, cf. [Bock 1987]

SCP and GGN do not converge to every local minimum. This can help to avoid "bad" local minima, as discussed next.



Regard maximum likelihood estimation problem $\boxed{\min_w \phi(M(w) - y)}$ with nonlinear model $M : \mathbb{R}^n \to \mathbb{R}^N$ and measurements $y \in \mathbb{R}^N$. Assume penalty ϕ is symmetric with $\phi(-z) = \phi(z)$ as is the case for symmetric error distributions. At a solution w^*, we can generate "mirror measurements" $y_{\text{mr}} := 2M(w^*) - y$ obtained by reflecting the residuals. From a statistical point of view, y_{mr} should be as likely as y.

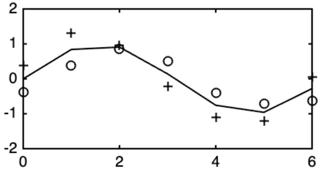
SCP Divergence $\Leftrightarrow$ Minimum unstable under mirroring



Theorem 2 [Diehl and Messerer 2019] generalizing [Bock 1987]

Regard a local minimizer w^* of $f(w) = \phi(M(w) - y)$ with $\nabla^2 f(w^*) \succ 0$. If the necessary SCP-GGN-convergence condition $B_{\mathrm{GGN}}(w^*) \succeq \frac{1}{2}\nabla^2 f(w^*)$ does not hold, then w^* is a stationary point of the mirror problem but **not** a local minimizer.

SCP Divergence $\Leftrightarrow$ Minimum unstable under mirroring

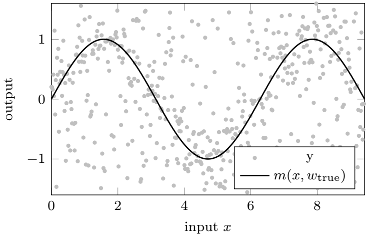


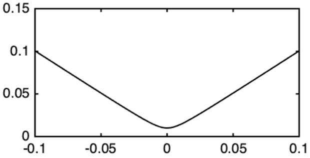
Theorem 2 [Diehl and Messerer 2019] generalizing [Bock 1987]

Regard a local minimizer w^* of $f(w) = \phi(M(w) - y)$ with $\nabla^2 f(w^*) \succ 0$. If the necessary SCP-GGN-convergence condition $B_{\mathrm{GGN}}(w^*) \succeq \frac{1}{2}\nabla^2 f(w^*)$ does not hold, then w^* is a stationary point of the mirror problem but **not** a local minimizer.

*Sketch of proof: use $M(w^*) - y_{\mathrm{mr}} = y - M(w^*)$ to show that
$\nabla f_{\mathrm{mr}}(w^*) = J(w^*)^\top (y - M(w^*)) = 0$ and
$\nabla^2 f_{\mathrm{mr}}(w^*) = B_{\mathrm{GGN}}(w^*) - E_{\mathrm{GGN}}(w^*) = 2B_{\mathrm{GGN}}(w^*) - \nabla^2 f(w^*) \not\succeq 0$

Illustrative example (experiments conducted by Florian Messerer)





Regard scalar model with unknown parameter w

$$m(x, w) := \sin(wx)$$

with input output measurements (x_i, y_i) for $i = 1, \dots, N$ (left plot).
Define $M_i(w) := m(x_i, w)$ and $F(w) := M(w) - y$
As outer convexity we use a Huber-like penalty (right plot)

$$\phi(z) := \frac{1}{N} \sum_{i=1}^{N} \sqrt{\delta^2 + z_i^2}, \qquad (3)$$

Implementation of SCP and GGN

- obtain all needed derivatives from CasADi via MATLAB interface
- use linear solve ("backslash") for GGN
- formulate SCP subproblem as second order cone program (SOCP):

$$\min_{\substack{w \in \mathbb{R}, \\ s \in \mathbb{R}^N}} \quad \sum_{i=1}^{N} s_i$$

$$\text{s.t.} \quad \sqrt{\delta^2 + F_{\mathrm{lin},i}(w; w_k)^2} \leq s_i \quad \text{for } i = 1, \ldots, N$$

- use Gurobi via CasADi as SOCP solver

CasADi for Optimization Modelling

CasADi

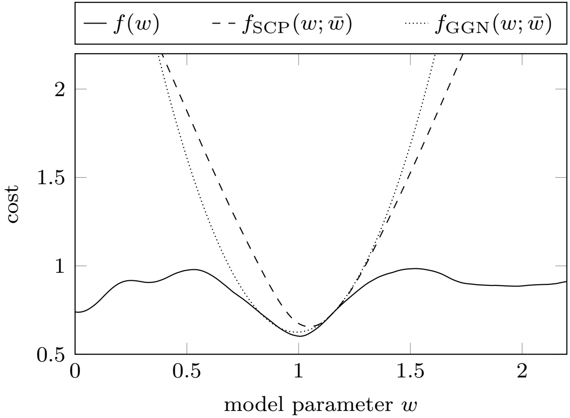
http://casadi.org

- A software framework for nonlinear optimization and optimal control
- "Write an efficient optimal control solver in a few lines"
- Implements automatic differentiation (AD) on sparse matrix-valued computational graphs in C++11
- Front-ends to Python, Matlab and Octave
- Supports C code generation
- Back-ends to SUNDIALS, CPLEX, Gurobi, qpOASES, IPOPT, KNITRO, SNOPT, SuperSCS, OSQP, ...
- Developed by Joel Andersson and Joris Gillis


Nov 18-20: Hands-on CasADi course on optimal control: http://hasselt2019.casadi.org

Jupyter CasADi demo for this talk: http://fgs.casadi.org

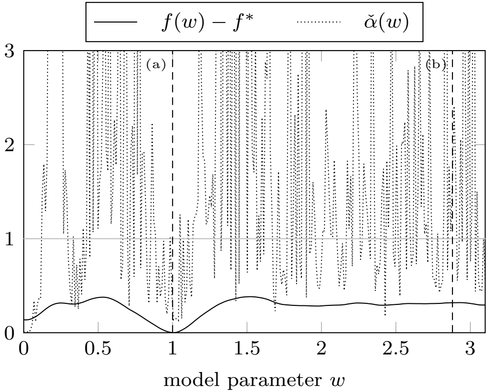
Visualization of SCP and GGN Subproblems



$(\bar{w} = 1.15)$

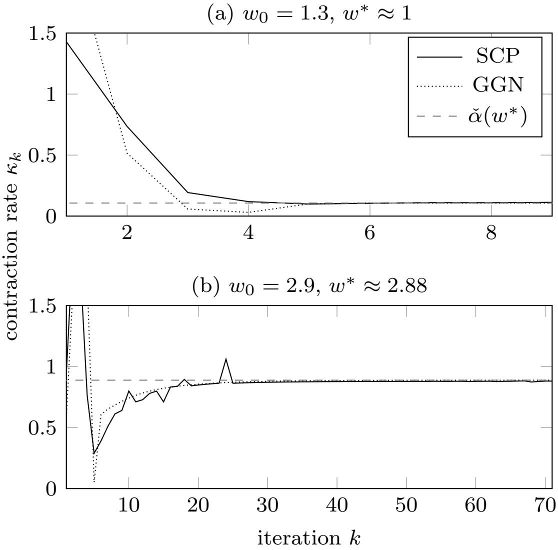
Objective and Local Contraction Rate

$f(w)$ and local rate $\check{\alpha}(w) = \frac{|\nabla^2 f(w) - B_{\mathrm{GGN}}(w)|}{|B_{\mathrm{GGN}}(w)|}$



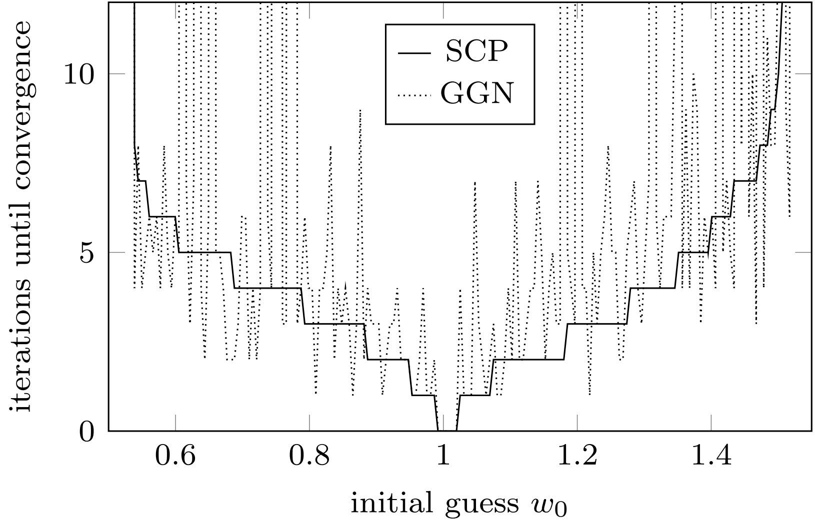
Fast contraction rate at global minimum $w^* \approx 1$, otherwise slower.

Empirical Contraction Rates in Agreement with Theorem 1



$$\kappa_k = \frac{|w_{k+1} - w_k|}{|w_k - w_{k-1}|}$$

Iteration count: SCP more predictable than GGN



Overview

- Smooth unconstrained problems
  - Sequential Convex Programming (SCP)
  - Generalized Gauss-Newton (GGN)
  - Local convergence analysis
  - Desirable divergence
  - Illustrative example in CasADi
- Constrained problems
  - Sequential Convex Programming (SCP)
  - Sequential Linear Programming (SLP)
  - Constrained Generalized Gauss-Newton (CGGN)
  - Sequential Convex Quadratic Programming (SCQP)
- Recent Progress in Dynamic Optimization and Applications

Sequential Convex Programming (SCP)

$$\begin{aligned}
\underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & \phi_0(F_0^{\text{lin}}(w; \bar{w})) \\
\text{subject to} \quad & F_i^{\text{lin}}(w; \bar{w}) \in \Omega_i, \quad i = 1, \dots, m, \\
& G^{\text{lin}}(w; \bar{w}) = 0
\end{aligned}$$

- obtain w_{k+1} as solution of convex problem at $\bar{w} = w_k$
- only primal variables $\bar{w}$ form "optimizer state", no multipliers needed
- SCP is affine invariant

A General Smooth NLP Formulation

From now on assume more smoothness and regard a smooth NLP

$$\begin{aligned} \underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & \underbrace{\phi_0(F_0(w))}_{=:f_0(w)} \\ \text{subject to} \quad & \underbrace{\phi_i(F_i(w))}_{=:f_i(w)} \leq 0, \quad i = 1, \ldots, m, \\ & G(w) = 0 \end{aligned}$$

with convex C^2 functions $\phi_0, \phi_1, \ldots, \phi_m$.

Sequential Linear Programming (SLP)

If functions $\phi_0, \ldots, \phi_m$ are linear, SCP just solves linear programs (LP):

$$\begin{aligned} \underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & f_0^{\text{lin}}(w; \bar{w}) \\ \text{subject to} \quad & f_i^{\text{lin}}(w; \bar{w}) \leq 0, \quad i = 1, \ldots, m, \\ & G^{\text{lin}}(w; \bar{w}) = 0 \end{aligned}$$

- might be called **Sequential Linear Programming (SLP)** ("Method of Approximation Programming" by Griffith & Stewart, 1961)
- equivalent to standard SQP with zero Hessian
- SLP only attracted to NLP solutions in vertices of feasible set
- works very well for L1-estimation [Bock, Kostina, Schlöder 2007]
- converges **quadratically** once correct active set is discovered

Picture from Griffith and Stewart's original 1961 paper

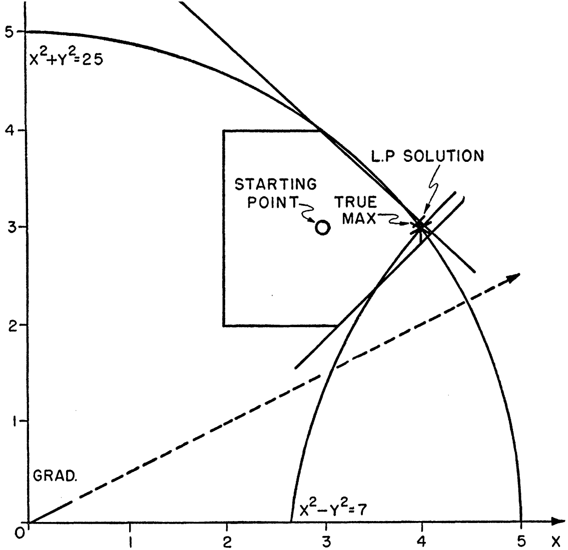


FIG. 4. Third LP problem

Constrained Generalized Gauss-Newton [Bock 1987]

Use $B_{\mathrm{GGN}}(w) := J_0(w)^\top \nabla^2 \phi_0(F_0(w)) J_0(w)$ and solve convex quadratic program (QP)

$$\begin{aligned} \underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & f_0^{\mathrm{lin}}(w;\bar{w}) + \frac{1}{2}(w-\bar{w})^\top B_{\mathrm{GGN}}(\bar{w})(w-\bar{w}) \\ \text{subject to} \quad & f_i^{\mathrm{lin}}(w;\bar{w}) \leq 0, \quad i = 1,\ldots,m, \\ & G^{\mathrm{lin}}(w;\bar{w}) = 0 \end{aligned}$$

- like SCP, the method is multiplier free
- also affine invariant

Remark: for least-squares objectives, this method is due to [Bock 1987]. In mathematical optimization papers, Bock's method is called "the Generalized Gauss-Newton (GGN) method" because it generalizes the GN method to constrained problems. To avoid a notation clash with computer science we prefer to call Bock's method "the Constrained Gauss-Newton (CGN) method".

Sequential Convex Quadratic Programming (SCQP) [Verschueren et al. 2016]

$$B_{\mathrm{SCQP}}(w,\mu) := J_0(w)^\top \nabla^2\phi_0(F_0(w))J_0(w) + \sum_{i=1}^{m} \mu_i J_i(w)^\top \nabla^2\phi_i(F_i(w))J_i(w)$$

$$\begin{aligned}
\underset{w \in \mathbb{R}^{n_w}}{\text{minimize}} \quad & f_0^{\text{lin}}(w;\bar{w}) + \frac{1}{2}(w-\bar{w})^\top B_{\mathrm{SCQP}}(\bar{w},\bar{\mu})(w-\bar{w}) \\
\text{subject to} \quad & f_i^{\text{lin}}(w;\bar{w}) \leq 0, \quad i = 1,\dots,m, \quad | \quad \mu^+, \\
& G^{\text{lin}}(w;\bar{w}) = 0
\end{aligned}$$

- again, only a QP needs to be solved in each iteration
- again, affine invariant
- "optimizer state" contains both, $\bar{w}$ and inequality multipliers $\bar{\mu}$
- $B_{\mathrm{SCQP}}(w,\mu) \succeq B_{\mathrm{GGN}}(w)$ i.e. more likely to converge than GGN
- SCQP has same contraction rate as SCP, characterized by two LMI on the reduced Hessian approximation [Messerer &D., 2019]

Overview

- Smooth unconstrained problems
  - Sequential Convex Programming (SCP)
  - Generalized Gauss-Newton (GGN)
  - Local convergence analysis
  - Desirable divergence
  - Illustrative example in CasADi
- Constrained problems
  - Sequential Convex Programming (SCP)
  - Sequential Linear Programming (SLP)
  - Constrained Generalized Gauss-Newton (CGGN)
  - Sequential Convex Quadratic Programming (SCQP)
- Recent Progress in Dynamic Optimization and Applications

Dynamic Optimization in a Nutshell

$$\begin{aligned} \underset{x,\,u}{\text{minimize}} \quad & \sum_{i=0}^{N} \varphi_i(F_i(x_i,u_i)) + \varphi_N(F_N(x_N)) \\ \text{subject to} \quad & x_{i+1} = S_i(x_i,u_i), \quad i = 0,\dots,N-1, \\ & H_i(x_i,u_i) \in \Omega_i, \qquad i = 0,\dots,N-1, \\ & H_N(x_N) \in \Omega_N \end{aligned}$$

- variables $w = (x, u)$ with $x = (x_0, \dots, x_N)$ and $u = (u_0, \dots, u_{N-1})$
- convexities in φ_i (e.g. quadratic) and Ω_i (e.g. polyhedral)
- nonlinearities in dynamic system S_i and constraint functions F_i, H_i
- often: S_i result of time integration (direct multiple shooting)

Recent Algorithmic Progress

- Inexact Newton with Iterated Sensitivities (INIS) - Rien Quirynen
- Generalized Nonlinear Static Feedback Integration - Jonathan Frey
- Zero Order Moving Horizon Estimation - Katrin Baumgärtner
- Advanced Step Real-Time Iteration - Armin Nurkanovic
- Convex Inner Approximation (CIAO) for mobile robots - Tobias Schöls

Recent Progress on Software and Applications

- **BLASFEO:** BLAS for Embedded Optimization - Gianluca Frison
- **acados:** real-time SCQP type algorithms for nonlinear model predictive control (NMPC) - Robin Verschueren, Dimitris Kouzoupis, Gianluca Frison, et al.

(both under permissive open-source BSD license)

Recent real-world NMPC applications:

- NMPC of small quadcopters in Freiburg - Barbara Barrros
- NMPC of human sized quadcopters in California - Andrea Zanelli
- 4 kHz NMPC of reluctance synchronous motor in Munich - Andrea Zanelli
- 100 Hz NMPC of Freiburg race cars - Daniel Kloeser
- 10 Hz CIAO-NMPC of mobile robots at Bosch - Tobias Schöls

Some Real-Time SCP Applications

Conclusions

- Discussed a family of Iterative Convex Approximation Methods
- Sequential Convex Programming (SCP) conceptually simplest ("linearize nonlinearities, keep all convex structures") and most widely applicable e.g. to nonlinear cone programming.
- Generalized Gauss-Newton (GGN) cheaper iterations as only QPs are solved. Unconstrained GGN and constrained SCQP show identical local convergence rates as SCP for smooth problems.
- Local contraction rate for both SCP and GGN equals smallest α satisfying $-\alpha B_{\mathrm{GGN}}(w^*) \preceq \nabla^2 f(w^*) - B_{\mathrm{GGN}}(w^*) \preceq \alpha B_{\mathrm{GGN}}(w^*)$
- Divergence might in fact be a desirable property in estimation as it avoids being attracted by statistically unstable local minima

Some References 1

- J. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. Mathematical Programming Computation, 2018.
- H.G. Bock, E. Kostina, and J.P. Schlöder. Numerical methods for parameter estimation in nonlinear differential algebraic equations. GAMM Mitteilungen, 30/2:376-408, 2007.
- J. Martens. Second-Order Optimization for Neural Networks. PhD thesis, University of Toronto, 2016.
- N. Schraudolph. Fast curvature matrix-vector products for second-order gradient descent. Neural Computation, 14(7):1723–1738, 2002.
- Q. Tran-Dinh. Sequential Convex Programming and Decomposition Approaches for Nonlinear Optimization. PhD thesis, KU Leuven, 2012.
- R. Verschueren, N. van Duijkeren, R. Quirynen, and M. Diehl. Exploiting convexity in direct optimal control: a sequential convex quadratic programming method. In Proceedings of the IEEE Conference on Decision and Control (CDC), 2016.

Some References 2

- M. Diehl and F. Messerer. Local Convergence of Generalized Gauss-Newton and Sequential Convex Programming, CDC, 2019 (accepted)
- N. Van Duijkeren. Online Motion Control in Virtual Corridors for Fast Robotic Systems, PhD thesis, KU Leuven, 2019.
- M. Diehl, F. Jarre, C. Vogelbusch. Loss of superlinear convergence for an SQP-type method with conic constraints, SIAM Journal on Optimization, 2006.
- Q. Tran-Dinh, S. Gumussoy, W. Michiels, M. Diehl. Combining convex-concave decompositions and linearization approaches for solving BMIs, with application to static output feedback, IEEE Transactions on Automatic Control, 2012.
- R.E. Griffith and R.A. Stewart. A Nonlinear Programming Technique for the Optimization of Continuous Processing Systems, Management Sci., 7: 379-392, 1961.