Operator Splitting Based Newton-type Method for Constrained Optimization

Titus Pinta, Russell Luke

Institut für Numerische und Angewandte Mathematik
Universität Göttingen

7 June 2024
Optimization Techniques in Quantum Chemistry
RWTH Aachen

Deutsche Forschungsgemeinschaft
German Research Foundation

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

inv opt KTG 2088 stat

Main Results

Given $f : \mathbb{R}^n \to \mathbb{R}$, C^1 with Newton differentiable gradient (e.g. piecewise $\mathcal{C}^2$), $G : \mathbb{R}^n \to \mathbb{R}^m$, differentiable and Newton solvable (e.g. $\mathcal{C}^1$)

Problem

Find

$$\operatorname*{argmin}_{y \in M := \{x \mid G(y) \leq 0\}} f(x)$$

where P_M is not easy to compute (e.g. quadratic constraints).

Result

The Newton splitting algorithm converges superlinearly.

Warm Up: Smooth Equality Constrains

Example: Newton on a Subspace

Given $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{C}^2$ and $S = \{y | Ay + b = 0\} \subseteq \mathbb{R}^n$

$$\underset{x \in S}{\operatorname{argmin}} \; f(x)$$

Solution:

$$\mathcal{N} : S \to \mathbb{R}^n \quad \mathcal{N}x = x - \nabla^2 f|_S(x)^{-1} \nabla f|_S(x)$$

where: $T = \{y \mid Ay = 0\}$
$\nabla f|_S(x) = \mathrm{P}_T \nabla f(x)$
$\nabla^2 f|_S(x) : T \to T \;\; \nabla^2 f|_S(x) = \mathrm{P}_T \circ \nabla^2 f(x)$

Can we extend this to general smooth equality constrains?

Newton Splitting

Model

$$\underset{y \in M := \{x \mid G(y)=0\}}{\operatorname{argmin}} f(x) \iff \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) + \iota_M(x)$$

First Order Optimality

$$0 \in \nabla f(x) + N_M(x) \iff 0 \in \nabla f|_{T_M(x)}(x) + \iota_M(x)$$

Splitting of the Model

$$\underbrace{0 \in \nabla f|_{T_M(x)}(x)}_{(\mathcal{P}_1)} \qquad \underbrace{0 \in \iota_M(x)}_{(\mathcal{P}_2)}$$

Idea: apply Newton's method to each problem subsequently

Newton Splitting contd.

Newton Step for $(\mathcal{P}_2)$

$G(x) = 0$ defines a set: $\mathcal{N}_{\mathcal{P}_2}$ should be a set

$$S(x) = \{y \mid G(x) + \nabla G(x)(y - x) = 0\}$$

Newton step to solve $G(x) = 0$

$$\mathcal{N}_{\mathcal{P}_2} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n, \quad \mathcal{N}_{\mathcal{P}_2} x := (\mathrm{P}_{S(x)}\, x,\ S(x))$$

In particular for $G : \mathbb{R}^n \to \mathbb{R}^n$ $\nabla G(x)$ invertible

$$\mathcal{N}_{\mathcal{P}_2} x = (x - \nabla G(x)^{-1} G(x),\ \{x - \nabla G(x)^{-1} G(x)\})$$

Newton Splitting contd.

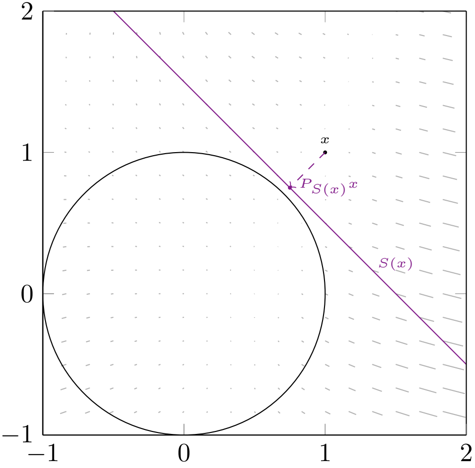

Figure: The $\mathcal{N}_{\mathcal{P}_2}$ step

Newton Splitting contd.

Recall

We can solve

$$\underset{S=\{y|Ay+b=0\}}{\operatorname{argmin}} \quad f(x)$$

Use the set produced by $\mathcal{N}_{\mathcal{P}_2}$ as S

Newton Step for $(\mathcal{P}_1)$

Newton step to solve $\nabla f|_{T_M(x)} = 0$ (i.e. $\nabla f|_{S(x)} = 0$)

$$\mathcal{N}_{\mathcal{P}_1} : \text{range}\left(\mathcal{N}_{\mathcal{P}_1}\right) \to \mathbb{R}^n, \quad \mathcal{N}_{\mathcal{P}_1}(x, S) = x - \nabla^2 f|_S(x)^{-1} \nabla f|_S(x)$$

Newton Splitting contd.

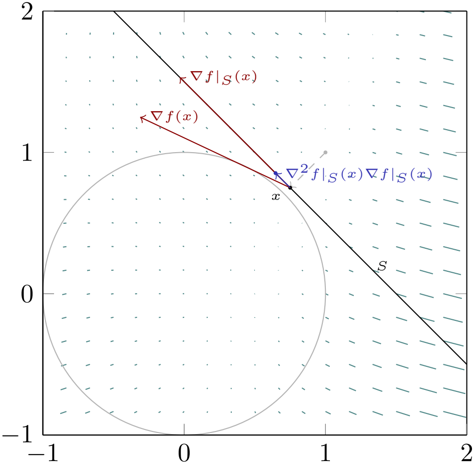

Figure: The $\mathcal{N}_{\mathcal{P}_1}$ step

Newton Splitting contd.

The Operator

$$\mathcal{N} = \mathcal{N}_{\mathcal{P}_1} \circ \mathcal{N}_{\mathcal{P}_2}$$

Putting it all together:

$$\mathcal{N}x = P_{S(x)}x - \nabla^2 f|_{S(x)}(P_{S(x)}x)^{-1} \nabla f|_{S(x)}(P_{S(x)}x)$$

Recall the definitions of $\nabla f|_{S(x)}$ and $\nabla^2 f_{S(x)}$

$T(x) = T_{S(x)}(x) = \{y \mid \nabla G(x) y = 0\}$
$\nabla f|_{S(x)}(x) = \mathrm{P}_{T(x)} \nabla f(x)$
$\nabla^2 f|_{S(x)}(x) : T(x) \to T(x) \quad \nabla^2 f|_{S(x)}(x) = P_{T(x)} \circ \nabla^2 f(x)$

$$\mathcal{N}x = P_{S(x)}x - P_{T(x)} \nabla^2 f(P_{S(x)}x)^{-1} P_{T(x)} \nabla f(P_{S(x)}x)$$

Newton Splitting Algorithm

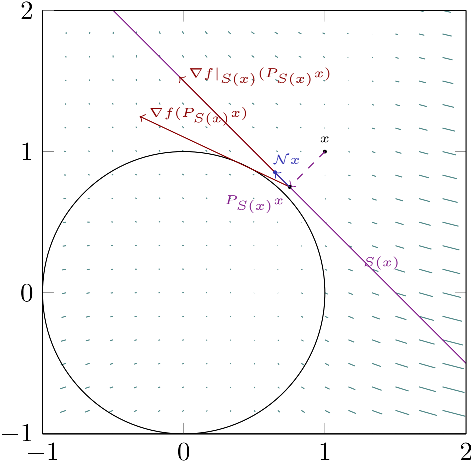

Figure: One Iteration of the Algorithm

Newton Splitting Algorithm

$$\mathcal{N}x^k = P_{S(x^k)}x^k - P_{T(x^k)}\nabla^2 f(P_{S(x^k)}x^k)^{-1} P_{T(x^k)}\nabla f(P_{S(x^k)}x^k)$$

Algorithm 1 Newton Splitting Algorithm

$k \leftarrow 0$
repeat
  $x^{k+1} = \mathcal{T}x^k$: specifically, compute
    $y = P_{S(x^k)}x^k$:
      $y \leftarrow \nabla G(x^k)^+ \nabla G(x^k) \nabla G(x^k)^+ (\nabla G(x^k) x^k - G(x^k))$
    $F = P_{S(x^k)}\nabla f(P_{S(x^k)}x^k)$:
      $F \leftarrow \nabla f(y) - \nabla G(x^k)^+ \nabla G(x^k) \nabla f(y)$
    $H = P_{S(x)}\nabla^2 f(P_{S(x)}x) P_{S(x)}$:
      $H \leftarrow (\mathrm{Id} - \nabla G(x^k)^+ \nabla G(x^k)) \nabla^2 f(y) (\mathrm{Id} - \nabla G(x^k)^+ \nabla G(x^k))$
    $x_{k+1} \leftarrow y - H^{-1}F$
until <convergence>

Inequality Constrains

Recap: Newton on a Subspace

Given $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{C}^2$ and $S = \{y | Ay + b = 0\} \subseteq \mathbb{R}^n$

$$\underset{x \in S}{\operatorname{argmin}} \; f(x)$$

Solution:

$$\mathcal{N} : S \to \mathbb{R}^n \quad \mathcal{N}x = x - \nabla^2 f|_S(x)^{-1} \nabla f|_S(x)$$

where: $T = \{y \mid Ay = 0\}$
$\nabla f|_S(x) = \mathsf{P}_T \nabla f(x)$
$\nabla^2 f|_S(x) : T \to T \;\; \nabla^2 f|_S(x) = \mathsf{P}_T \circ \nabla^2 f(x)$

Inequality Constrains

Given $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{C}^2$ and $S = \{y | Ay + b \leq 0\} \subseteq \mathbb{R}^n$

$$\underset{x \in S}{\operatorname{argmin}} \; f(x)$$

Solution:

$$\mathcal{N} : S \to \mathbb{R}^n \quad \mathcal{N}x = x - \nabla^2 f|_S(x)^{-1} \nabla f|_S(x)$$

where: $T(x) = \{y \mid \exists t_n \downarrow 0, \exists x_n \in S(x) \to x : (x_n - x)/t_n \to y\}$
$\nabla f|_S(x) = \mathrm{P}_{T(x)} \nabla f(x)$
$\nabla^2 f|_S(x) : T(x) \to T(x) \; \nabla^2 f|_S(x) = \mathrm{P}_{T(x)} \circ \nabla^2 f(x)$
$\nabla^2 f|_S(x)$ is a positive-linear map:

$$\nabla^2 f|_S(x)^{-1}(v_1 + v_2) = \nabla^2 f|_S(x)^{-1} v_1 + \nabla^2 f|_S(x)^{-1} v_2$$

Newton Splitting

$$\underbrace{0 \in \nabla f|_{T_M(x)}(x)}_{(\mathcal{P}_1)} \qquad \underbrace{0 \in \iota_M(x)}_{(\mathcal{P}_2)}$$

Newton Step for $G(x) \leq 0$

$$\mathcal{N}_{\mathcal{P}_2} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n, \quad \mathcal{N}_{\mathcal{P}_2} x := (\mathrm{P}_{S(x)}\, x,\; S(x))$$

where

$$S(x) = \{y \mid G(x) + \nabla G(x)(y - x) \leq 0\}$$

Newton Step for $\nabla f|_{T_M(x)} = 0$

$$\mathcal{N}_{(\mathcal{P}_1)} : \text{range}\left(\mathcal{N}_{(\mathcal{P}_2)}\right) \to \mathbb{R}^n, \quad \mathcal{N}_{(\mathcal{P}_2)}(x, S) = x - \nabla^2 f|_S(x)^{-1} \nabla f|_S(x)$$

Analysis Warm Up: Nonsmooth Unconstrained

Newton Differentiability

Definition

Given $F : \mathbb{R}^n \to \mathbb{R}^n$ and $y \in \mathbb{R}^n$, F is pointwise Newton differentiable at y if there exists $\mathbb{H} : \mathbb{R} \rightrightarrows \mathbb{R}^{n \times n}$ and

$$\exists \gamma \in [0, \infty] : \|F(x) - F(y) - H(x)(x - y)\| \leq O\left(\|x - y\|^{1+\gamma}\right)$$

holds for all $H(x) \in \mathbb{H}(x)$ for all x.

Convergence: Classical Newton's Method

Given F pointwise Newton differentiable at $\bar{x}$, isolated with $F(\bar{x}) = 0$, blood$H(x)^{-1}$ bounded for all $H(x) \in \mathbb{H}(x)$ for all x near $\bar{x}$, the sequence

$$x^{k+1} := Tx^k = x^k - H(x^k)^{-1} F(x^k)$$

is convergent with rate $1 + \gamma$ (superlinear if $\gamma > 0$) to $\bar{x}$ for all x^0 close enough to $\bar{x}$

Examples

1. f strong convex with L Lipschitz gradient $\Longrightarrow \nabla f$ Newton diff. with $H = L\,\mathrm{Id}$
2. $f \in \mathcal{C}^2 \Longrightarrow \nabla f$ Newton diff. with $H = \nabla^2 f$
3. ∇f γ-semismooth $\Longrightarrow \nabla f$ Newton diff. with $H = \partial \nabla f$

∇f γ-semismooth if $\partial \nabla f$ exists and for all x, y and $C(x) \in \partial \nabla f(x)$

$$\|\nabla f(x) - \nabla f(y) - C(x)(x-y)\| \leq O\left(\|x-y\|^{1+\gamma}\right)$$

where

$$\partial \nabla f(x) = \overline{\mathrm{conv}}\{ \lim_{\substack{x_n \to x \\ \exists \nabla^2 f(x_n)}} \nabla^2 f(x_n)\}$$

Analysis: Main Result

Constrained Newton Differentiability

Recap

$S(x) = \{y \mid G(x) + \nabla G(x)(y - x) \leq 0\}$
$T(x) = \{y \mid \exists t > 0 : P_{S(x)}x + ty \in S(x)\}$ and
$v \in \mathbb{R}^n \quad v|_{S(x)} = P_{T(x)}v$

Definition

F pointwise Newton differentiable at $\bar{x}$, $F_{S(\bar{x})}(\bar{x}) = 0$, with respect to the differentiable constrains G there is $H : \mathbb{R}^n \to \mathbb{R}^{n\times n}$

$\exists\gamma \in [0, 1]$:

$$\|F|_{S(x)}(P_{S(x)}x)) - H|_{S(x)}(P_{S(x)}x)(P_{T(x)}(x - \bar{x}))\| \leq O\left(\|x - \bar{x}\|^{1+\gamma}\right)$$

Newton Solvability

Definition

$G : \mathbb{R}^n \to \mathbb{R}^m$ is Newton Solvable at $\bar{x}$ if

$$\|P_{S(x)}\bar{x} - \bar{x}\| \le O\left(\|x - \bar{x}\|^{1+\gamma}\right)$$

Intuition

$G : \mathbb{R}^n \to \mathbb{R}^n$, for all x, $G(x) \ge 0$, $\exists! \bar{x}$ with $G(x) = 0$

$$P_{S(x)}\bar{x} = P_{\{y \mid G(x) + \nabla G(x)(y-x)\}} = \mathcal{N}_C x = x - \nabla G(x)^{-1} G(x)$$

Examples

1. $G : \mathbb{R}^n \to \mathbb{R}^n$, for all x, $G(x) \ge 0$, $\exists! \bar{x}$ with $G(x) = 0$
   1.1 $G \in \mathcal{C}^1$ and $\nabla G(\bar{x})$ invertible
   1.2 G semismooth and $\partial G(x)$ for all x near $\bar{x}$ invertible
2. $G : \mathbb{R}^n \to \mathbb{R}^m$, for all x G Newton differentiable, for all x, $G(x) \ge 0$ and $\nabla G(x)^+$ bounded

Examples

Smooth Functions

$f \in \mathcal{C}^2$, G Newton solvable, then f Newton differentiable at $\bar{x}$ with respect to G with $h = \nabla^2 f$
If $\nabla^2 f(\bar{x})$ invertible, then $\nabla^2 f(x)|_{S(x)}$ invertible

Semismooth Functions

∇f γ semismooth, G Newton solvable then f Newton differentiable at $\bar{x}$ with respect to G with $H = \partial \nabla f$
$\partial \nabla f|_{S(x)}(x)$ invertible required

Main Result

Problem

$$\operatorname*{argmin}_{y \in M := \{x \mid G(y) \leq 0\}} f(x) \Longrightarrow 0 \in \{y \mid G(y) \leq 0\} \cap \nabla f|_{T_M(x)}(x)$$

Convergence

Given F pointwise Newton differentiable at $\bar{x}$ with respect to the Newton solvable constrains G, $S(x)$ and $T(x)$ defined as above, $F|_{S(\bar{x})}(\bar{x}) = 0$ isolated and $H|_{S(x)}(x)^{-1}$ bounded

$$\begin{aligned} \mathcal{N}x &= P_{S(x)}x - H|_{S(x)}(P_{S(x)}x)^{-1}F|_{S(x)}(P_{S(x)}x) \\ &= P_{S(x)} - (P_{T(x)}H(P_{S(x)}x)P_{T(x)})^{-1}P_{T(x)}F(P_{S(x)}x) \end{aligned}$$

superlinearly convergent with rate $1 + \gamma$ (superlinear if $\gamma > 0$)

Numerical Experiments

$$\min_{x\in\mathbb{R}^n} \sum_i (|\mathcal{F}\tilde{x}_i| - b_i), \quad \text{where } \tilde{x} = (0, x, 0) \geq 0$$

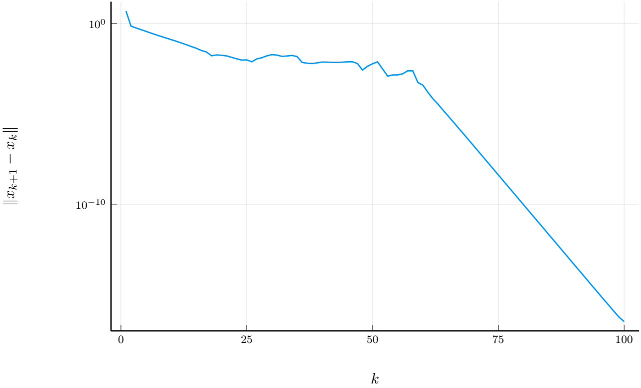

Figure: Step length for phase retrival

Context

SQP

Sequential Quadratic Programming (1960s)

$$\Delta(x) = \underset{d \in S(x)}{\operatorname{argmin}} \frac{1}{2} d^T \nabla^2 \mathcal{L} f(x) d + \nabla f(x)^T d$$

$$Tx = x + \Delta(x)$$

Newton Splitting

$$\Delta(x) = \underset{d \in T(x)}{\operatorname{argmin}} \frac{1}{2} d^T P_{T(x)} \nabla^2 f(P_{S(x)} x) d + P_{T(x)} \nabla f(P_{S(x)} x)^T d$$

$$Tx = P_{S(x)} x + \Delta(x)$$

Main Results again

Given $f : \mathbb{R}^n \to \mathbb{R}$, C^1 with Newton differentiable gradient (e.g. piecewise $\mathcal{C}^2$), $G : \mathbb{R}^n \to \mathbb{R}^m$, differentiable and Newton solvable (e.g. $\mathcal{C}^1$)

Problem

Find

$$\operatorname*{argmin}_{y \in M := \{x \mid G(y) \leq 0\}} f(x)$$

where P_M is not easy to compute (e.g. quadratic constraints).

Result

The Newton splitting algorithm converges superlinearly.

References

[1] T. Pinta
Operator Splitting Based Newton-type Method for Constrained Optimization.
work in progress

[2] Brokate, Martin and Ulbrich, Michael.
Newton Differentiability of Convex Functions in Normed Spaces and of a Class of Operators.
SIAM J. on Optimization, 32(2):1265-1287, January 2022.

[3] D. Kraft.
A software package for sequential quadratic programming.
Technical Report DFVLR-FB 88-28, Institut für Dynamik der Flugsysteme, Oberpfaffenhofen, July 1988.

[4] Liqun Qi and Jie Sun.
A Nonsmooth Version of Newton's Method.
Mathematical Programming, 58(1):353–367, January 1993.

Appendix 1 - Gradient Descent

From strong convexity:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq O(\|x - y\|^2)$$

So

$$\begin{aligned}
&\|\nabla f(x) - \nabla f(y) - L(x - y)\|^2 \\
&\quad = \|\nabla f(x) - \nabla f(y)\|^2 + L^2\|x - y\|^2 - L\langle \nabla f(x) - \nabla f(y), x - y \rangle \\
&\quad \leq L^2\|x - y\|^2 + L^2\|x - y\|^2 - L\langle \nabla f(x) - \nabla f(y), x - y \rangle \\
&\quad \leq O(\|x - y\|^2)
\end{aligned}$$

$$\|\nabla f(x) - \nabla f(y) - L(x - y)\| \leq O(\|x - y\|)$$

Appendix 2 - Main Proof

$$
\begin{aligned}
\|\mathcal{N}x - \bar{x}\| &= \|P_{S(x)}x - H|_{S(x)}(P_{S(x)}x)^{-1}F|_{S(x)}(P_{S(x)}x) - \bar{x}\| \\
&= \|H|_{S(x)}(P_{S(x)}x)^{-1}(H|_{S(x)}(P_{S(x)}x)P_{S(x)}x \\
&- F|_{S(x)}(P_{S(x)}x)) - \bar{x}\| \\
&= \|H|_{S(x)}(P_{S(x)}x)^{-1}(H|_{S(x)}(P_{S(x)}x)(P_{S(x)}x - P_{S(x)}\bar{x}) \\
&- F|_{S(x)}(P_{S(x)}x)) + P_{S(x)}\bar{x} - \bar{x}\| \\
&\leq \|H|_{S(x)}(P_{S(x)}x)^{-1}\|\|(H|_{S(x)}(P_{S(x)}x)(P_{S(x)}x - P_{S(x)}\bar{x}) \\
&- F|_{S(x)}(P_{S(x)}x))\| + \|P_{S(x)}\bar{x} - \bar{x}\| \\
&= \|H|_{S(x)}(P_{S(x)}x)^{-1}\|\|(H|_{S(x)}(P_{S(x)}x)(P_{T(x)}(x - \bar{x})) \\
&- F|_{S(x)}(P_{S(x)}x))\| + \|P_{S(x)}\bar{x} - \bar{x}\| \\
&\leq MO(\|x - \bar{x}\|^{1+\gamma}) + O(\|x - \bar{x}\|^{1+\gamma}) \\
&= O(\|x - \bar{x}\|^{1+\gamma})
\end{aligned}
$$

Appendix 3 - Semismooth and Newton Diff.

$$
\begin{aligned}
&\|P_{T(x)}\nabla f(P_{S(x)}x) - P_{T(x)}\nabla^2 f(P_{S(x)}x)(P_{S(x)}x - P_{S(x)}\bar{x})\| \\
&\|P_{T(x)}\nabla f(P_{S(x)}x) - P_{T(x)}\nabla f(P_{S(x)}\bar{x}) \\
&- P_{T(x)}\nabla^2 f(P_{S(x)}x)(P_{S(x)}x - P_{S(x)}\bar{x}) + P_{T(x)}\nabla f(P_{S(x)}\bar{x})\| \\
&\leq \|\nabla f(P_{S(x)}x) - \nabla^2 f(P_{S(x)})(P_{S(x)}x - P_{S(x)}\bar{x})\| + \|P_{T(x)}\nabla f(P_{S(x)}\bar{x})\| \\
&\leq O(\|P_{S(x)}\|\|x - \bar{x}\|^{1-\gamma}) + \|P_{T(x)}\nabla f(P_{S(x)}\bar{x})\| \\
&\leq O(\|P_{S(x)}\|\|x - \bar{x}\|^{1-\gamma}) + \|P_{T(x)}\nabla f(P_{S(x)}\bar{x}) - P_{L(\bar{x})}\nabla f(P_{S(\bar{x})}\bar{x})\| \\
&\leq O(\|P_{S(x)}\|\|x - \bar{x}\|^{1-\gamma}) + O(\|P_{S(x)}\bar{x} - \bar{x}\|)
\end{aligned}
$$