

Sparse Optimization with Applications to Cancer Research

BORIS MORDUKHOVICH

Department of Mathematics and Institute for AI and Data Science,
Wayne State University USA, <http://math.wayne.edu/boris>

lecture given at

Workshop on Optimization and Machine Learning

9th German-Polish Conference on Optimization

Białystok, Poland

September 2026

Overview

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

Proton Arc Therapy

Proton therapy is an advanced type of radiation treatment for cancer that uses high-energy protons to precisely kill tumor, while minimizing damage to surrounding healthy tissues. In proton therapy, especially [proton arc therapy](#), treatment delivery efficiency plays a crucial role in the routine clinical implementation. Previous investigations indicated that the number of energy layers and spot impact not only affects plan quality but also the treatment delivery time. Therefore, finding an [optimal sparsity level](#) of the energy layer or spot to ensure fast treatment delivery while maintaining clinically acceptable plan quality remains a significant challenge.

In recent years, the [\$\ell_0\$ -norm function](#), which is [nonsmooth and nonconvex](#), has gained significant attention in the field of medical physics. Using the ℓ_0 -norm encourages sparsity and enables the extraction of meaningful information from the noisy one and/or reducing the unnecessary delivery, which plays a pivotal role to ensure accurate diagnostics and patient treatment efficiency.

Proton Arc Therapy

Proton arc therapy delivers proton beams while the treatment machine rotates around the patient. A challenge to proton arc therapy is how to **reduce beam delivery time**. One approach is to decrease spot number. The alternating direction method of multipliers was applied in ^{1,2, 3} from the viewpoint of **ℓ_0 -sparsity optimization** in proton arc therapy to reduce the corresponding spot number. These strategies prioritize the principle of sparsity while assuming that optimal solutions are either nearly sparse, or can be made sparse through an appropriate transformation.

¹L. Zhao, J. You, G. Liu, X. Gang, X. Lu, and X. Ding, A novel simultaneous plan quality and beam delivery time SPARC optimization platform using alternating direction method of multipliers (ADMM), *Particle Therapy Cooperative Group*, 2022

²L. Zhao et al., The first direct method of spot sparsity optimization for proton arc therapy, *Acta Oncol. Stockh. Swed.* **62** (2023), 48–52

³Q. Fan, L. Zhao, X. Ding et. al, Optimizing spot-scanning proton arc therapy with a novel spot sparsity approach, *Medical Physics*, 2024

Main Problems

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

ℓ_0 -Norm function

The ℓ_0 -norm function $\|\cdot\|_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by:

$$\|x\|_0 = \text{number of nonzero components.}$$

Some examples in the $\mathbb{R}^3$ space:

$$\|(0, -1, 4)\|_0 = 2,$$

$$\|(1, 0, 0)\|_0 = 1,$$

$$\|(0, 0, 0)\|_0 = 0,$$

$$\|(1, 2, 3)\|_0 = 3.$$

ℓ_0 -Norm Function

We can interpret $\|\cdot\|_0$ as a **piecewise constant function** (e.g., in $\mathbb{R}^2$) by

$$f(x, y) = \begin{cases} 0 & \text{if } x = y = 0, \\ 1 & \text{if } x \neq 0, y = 0 \text{ or if } x = 0, y \neq 0, \\ 2 & \text{otherwise.} \end{cases} \quad (1)$$

The ℓ_0 -norm function is **nonsmooth**, **nonconvex** but **variationally convex**

Based on (1), the graph of the ℓ_0 -norm function in $\mathbb{R}^2$ is

$$\text{gph} \|\cdot\|_0 = \{(0, 0, 0)\} \cup \{(x, 0, 1) \mid x \neq 0\} \\ \cup \{(0, y, 1) \mid y \neq 0\} \cup \{(x, y, 2) \mid x, y \neq 0\}$$

ℓ_0 -Norm Function Graph

which is depicted in Figure 1

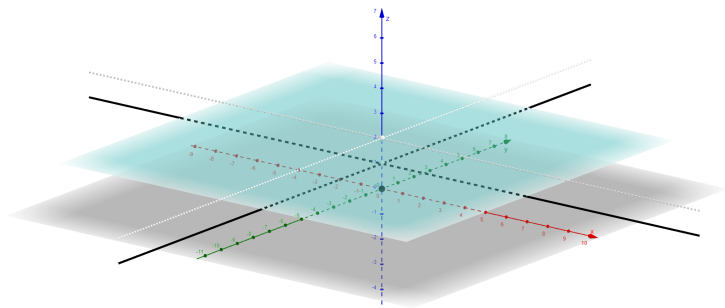


Figure: Graph of the ℓ_0 -norm function

Multiobjective ℓ_0 -Norm Optimization

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $F(x) := \{F_1(x), \dots, F_m(x)\}$, where

$$F_i(x) := f_i(x) + \|x\|_0$$

are such that f_i , for $i = \overline{1, m}$, are convex differentiable functions. The *multiobjective/vector ℓ_0 -norm optimization problem* is formulated as

$$(\mathbf{MO}\text{-}\ell_0\mathbf{NOP}) \quad \min_{x \in \mathbb{R}^n} F(x) = (F_1(x), \dots, F_m(x))$$

where the *optimal solution/vector minimizer* $\bar{x}$ pf F is understood in the sense of **Pareto efficiency**, i.e., there is no point x satisfying

$$F_i(x) \leq F_i(\bar{x}) \quad \text{for all } i = 1, \dots, m$$

with $F_{i_0}(x) < F_{i_0}(\bar{x})$ for some $i_0 \in \{1, \dots, m\}$

Single-Objective ℓ_0 -Norm Optimization

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex differentiable function whose gradient is *Lipschitz continuous* with a constant L , i.e.,

$$|f(x_1) - f(x_2)| \leq L\|x_1 - x_2\| \quad \text{for all } x_1, x_2 \in \mathbb{R}^n.$$

The **single-objective/scalar ℓ_0 -norm optimization problem** is formulated as

$$(\ell_0\text{NOP}) \quad \min_{x \in \mathbb{R}^n} f(x) + \|x\|_0$$

Example. Consider the quadratic function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) := (x - \frac{1}{2})^2 + \frac{1}{4}(y + 1)^2$. The graph of $f(x, y) + \|(x, y)\|_0$ is depicted in Figure 2.

Single-Objective ℓ_0 -Norm Optimization

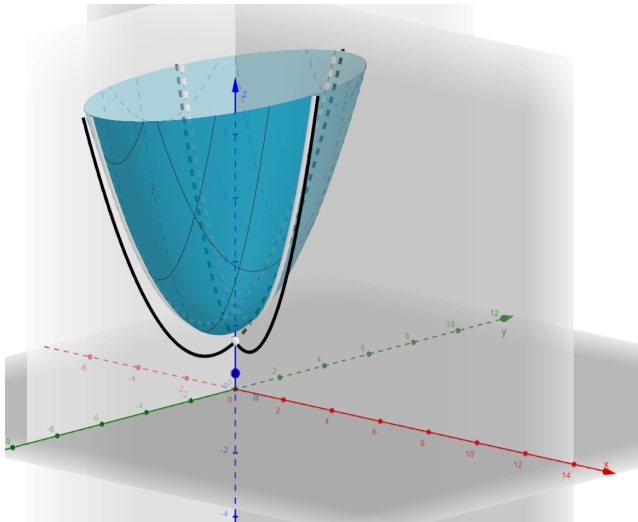


Figure: Graph of $f(\cdot) + \|\cdot\|_0$

Overview

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

Variational Analysis

Variational analysis has been recognized as an advanced and rapidly developed area of applied mathematics and operations research, which develops and implements powerful theoretical tools and numerical algorithms to solve large-scale problems arising in optimization, systems control, AI, data science, etc.

A characteristic feature of modern optimization problems and variational techniques is the unavoidable presence of *nondifferentiable* functions. This requires developing tools of **generalized differentiation** satisfying comprehensive **calculus rules**; see, e.g., the books ⁴, ⁵, ⁶ for the constructions used in this talk.

⁴R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, 1998

⁵B. S. Mordukhovich, *Variational Analysis and Applications*, 2018

⁶B. S. Mordukhovich, *Second-Order Variational Analysis in Optimization, Variational Stability, and Control: Theory, Algorithms, Applications*, 2024

Subgradients and Normals

Let $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ be an **extended-real-valued** function. The **subdifferential** of f at $\bar{x} \in \text{dom } f$ is

$$\partial f(\bar{x}) = \left\{ v \in \mathbb{R}^n \mid \begin{array}{l} \exists x_k \rightarrow \bar{x}, v_k \rightarrow v \text{ with } f(x_k) \rightarrow f(\bar{x}), \\ \liminf_{x \rightarrow x_k} \frac{f(x) - f(x_k) - \langle v_k, x - x_k \rangle}{\|x - \bar{x}\|} \geq 0 \end{array} \right\}$$

For convex functions, ∂f reduces to the subdifferential of convex analysis, while for nonconvex functions $\partial f(\bar{x})$ is generally nonconvex, e.g., for $f(x) = -|x|$ we have $\partial f(0) = \{-1, 1\}$. Nevertheless, ∂f enjoys **full calculus** based on **variational/extremal principles** of variational analysis

The **normal cone** $N_{\Omega}(\bar{x})$ to sets is defined via the subdifferential of the set indicator function

Coderivative and Second-Order Subdifferential

Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping.

The **coderivative** of F at $(\bar{x}, \bar{y}) \in \text{gph} F$ is defined by

$$D^*F(\bar{x}, \bar{y})(v) := \{ u \in \mathbb{R}^n \mid (u, -v) \in N_{\text{gph} F}(\bar{x}, \bar{y}) \}$$

Given $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ and $\bar{v} \in \partial f(\bar{x})$, the **second-order subdifferential/generalized Hessian** of f at $\bar{x}$ for $\bar{v}$ is

$$\partial^2 f(\bar{x}, \bar{v}) := (D^* \partial f)(\bar{x}, \bar{v})$$

Full calculus and computation are available for major classes of extended-real-valued functions and set-valued mappings ⁷

⁷B. S. Mordukhovich, *Second-Order Variational Analysis in Optimization, Variational Stability, and Control: Theory, Algorithms, Applications*, 2024.

Gerstewitz function

Given a closed convex cone $C \subset \mathbb{R}^m$, a direction $0 \neq k^0 \in C \setminus (-C)$, and a nonempty set $A \subset \mathbb{R}^m$ with $A - \mathbb{R}_+ k^0 \subset A$, the **Gerstewitz function**⁸ is

$$\phi_{A,k^0} := \inf\{t \in \mathbb{R} \mid y \in tk^0 + A\}$$

ϕ_A is a **convex function** if and only if A is a **convex set**. In this case, the convex subdifferential of ϕ_A at $\bar{y}$ is calculated by

$$\partial\phi_A(\bar{y}) = \{y^* \in A \mid \langle k^0, y^* \rangle = 1, \langle \bar{y}, y^* \rangle - \phi_A(\bar{y}) \geq \langle y, y^* \rangle, \quad y \in A\}.$$

⁸A. A. Khan, C. Tammer, and C. Zalinescu, *Set-Valued Optimization: An Introduction*, 2015

Scalarization for MO-L0NOP

Consider the multiobjective ℓ_0 -norm optimization problem. Then the **Gerstewitz scalarization** is

$$\phi_A(f_1(x) + \|x\|_0, \dots, f_m(x) + \|x\|_0).$$

Choosing $k^0 = (1, \dots, 1) \in \mathbb{R}^m$ and $\|x\|_0 = \lambda$, we can rewrite the Gerstewitz secularization as

$$\begin{aligned}\phi_A(f_1(x) + \|x\|_0, \dots, f_m(x) + \|x\|_0) &= \phi_A((f_1(x), \dots, f_m(x)) + \lambda k^0) \\ &= \phi_A(f_1(x), \dots, f_m(x)) + \lambda \\ &= \phi_A(f_1(x), \dots, f_m(x)) + \|x\|_0.\end{aligned}$$

Hence **MO- ℓ_0 NOP** is converted to the **single-objective problem**

$$\min \phi_A(f_1(x), \dots, f_m(x)) + \|x\|_0$$

Scalarization for MO-L0NOP

Another approach to scalarization of multiobjective problems is using the classical **Weighted Sum method**. We convert **MO- ℓ_0 NP** into a single-objective optimization problem by assigning a weight to each objective function and summing them. This can be written as

$$\textbf{Weight sum}(x) = \sum_{i=1}^m \omega_i f_i(x) + \|x\|_0$$

where $\omega_i \geq 0$, $\sum_{i=1}^m \omega_i = 1$. Note that the **Weight sum** can be viewed as a particular case of the Gerstewitz scalarization.

First-Order and Second-Order Algorithms

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

First-Order: Subgradient Type

$$x_{k+1} = x_k - \alpha_k v_k, \quad v_k \in \partial f(x_k)$$

Second-Order: Coderivative-Based Newton⁹

$$x_{k+1} = x_k + d_k, \quad -v_k \in \partial^2 f(x_k, v_k)(d_k), \quad v_k \in \partial f(x_k)$$

with using the **general** first-order and second-order subdifferentials

The **crucial issue** is the these subdifferentials are **explicitly computed** for the ℓ_0 -norm and scalarization functions via the given data

⁹P. D. Khanh, B. S. Mordukhovich and T. V. Phat, Coderivative-based Newton methods in structured nonconvex and nonsmooth optimization, Math. Program., DOI:10.1007/s10107-026-02420-7

Projection Mapping and Hyperplanes

Based on the first-order and second-order [subdifferential calculations](#) for ℓ_0 -norm, define the **projection mapping** $\tilde{l}_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ as follows: for $x_i = (x_{1,i}, \dots, x_{n,i}) \in \mathbb{R}^n$, let $\tilde{l}_i$ be an $n \times n$ matrix such that

$$\begin{aligned} a_{j,k} &= 0 && \text{if } j \neq k, \\ a_{j,j} &= 0 && \text{if } x_{j,i} = 0, \\ a_{j,j} &= 1 && \text{if } x_{j,i} \neq 0. \end{aligned}$$

Denote by $\tilde{\nabla} f$ and $\tilde{\nabla}^2 f$ the gradient and Hessian respect to inactive components x_j by: if $x_j = 0$ then $\frac{\partial f}{\partial x_j} = 0$, and $\frac{\partial^2 f}{\partial x_j x_i} = \frac{\partial^2 f}{\partial x_i x_j} = 0$ for all i . The obtained subdifferential calculation allow us to proceed with the [reduction to hyperplanes](#)

Projection Mapping and Hyperplanes

Claim 4.1.

The projection mapping $\tilde{l}_i$ is defined by the value of x_i , and clearly $\tilde{l}_i x_i = x_i$. For each projection mapping, we have the *hyperplane* $l_i = \tilde{l}_i(x_1, \dots, x_n)$

Theorem 4.1.

If $\bar{x}$ solves the **projection problem**

$$(\text{PP}) \quad \min_{x \in l_i} f(x),$$

then $\bar{x}$ is a local minimizer of $(\ell_0 \text{NOP})$

Subgradient-Based Algorithm for Single-Objective ℓ_0 -Optimization

Algorithm 1 Constant stepsize

Require: Given x_0 as start point, $\epsilon > 0$, step size $t < \frac{1}{L}$

if $\|\tilde{l}_k \nabla f(x_k)\| \geq \epsilon$ **then**
 $x_{k+1} = x_k - t \cdot \tilde{l}_k \nabla f(x_k)$
else
 Stop
end if

Subgradient-Based Algorithms for Single-Objective ℓ_0 -Optimization

Algorithm 2 Constant stepsize, more local minimizers

Require: Given x_0 as start point, $\epsilon > 0$, step size $t < \frac{1}{L}$

```
if  $\|\nabla f(x_k)\| \geq \epsilon$  then
  if  $\|\tilde{l}_k \nabla f(x_k)\| \geq \epsilon$  then
     $x_{k+1} = x_k - t \cdot \tilde{l}_k \nabla f(x_k)$ 
  else
     $x_{k+1} = x_k - t \cdot (I - \tilde{l}_k) \nabla f(x_k)$ 
  end if
else Stop
end if
```

Subgradient-Based Algorithm for Multiobjective ℓ_0 Optimization

Requirement: f_i have Lipschitz continuous gradients with modulus L_i for $i = \overline{1, m}$.

Algorithm 3 Constant stepsize, multiobjective

Require: Given x_0 as a starting point, $\epsilon > 0$, stepsize $t < \frac{1}{L}$ where $L := \max_{i=\overline{1, m}} L_i$, choose ω_i
if $\|\tilde{l}_k \nabla \text{Weight sum}(x_k)\| \geq \epsilon$ **then**
 $x_{k+1} = x_k - t \cdot \tilde{l}_k \nabla \text{Weight sum}(x_k)$
else Stop
end if

Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization

Algorithm 4 Subgradient-based method for Gerstewitz scalarization

Requirement: f_i are Lipschitz continuous with modulus L_i for $i = \overline{1, m}$.

Require: Given x_0 as start point, $\epsilon > 0$, step size t , $A \subset \mathbb{R}^m$ is a convex set with $A - \mathbb{R}_+^m \subset A$, $k^0 = (1, \dots, 1)$.

if $\|\phi_A(f_1, \dots, f_m)(x^{k+1}) - \phi_A(f_1, \dots, f_m)(x^k)\| \geq \epsilon$ or $k < 1000$ **then**

$x_{k+1} = x_k - t \cdot \tilde{l}_k J_f(x^k)^\top g^k$, where $J_f(x^k)$ is a Jacobian matrix at x^k and $g^k \in \partial \phi_A(f_1(x^k), \dots, f_m(x^k))$

$f_{best} = \min\{f_{best}, f(x^{k+1})\}$

else Stop

end if

Coderivative-Based Newton Algorithm for Single-Objective ℓ_0 Optimization

Algorithm 5 Local Newton-Type Method for ℓ_0 -Norm Optimization

Require: Given x_0 as starting point, $\epsilon > 0$

for $k = 0, 1, \dots$,

if $\|\tilde{\nabla} f(x^k)\| \geq \epsilon$ **then**,

Newton step: Choose $d^k \in \mathbb{R}^n$ satisfying

$$-\tilde{\nabla} f(x^k) = \tilde{\nabla}^2(x^k)(d^k)$$

Set $x^{k+1} := x^k + d^k$

else Stop

end if

Globalized Coderivative-Based Newtonian Algorithms for Single-Objective ℓ_0 Optimization

Algorithm 6 Armijo Backtracking Line Search Newton with Modification for ℓ_0 -Norm Optimization (1/2)

Require: $x_0, \epsilon > 0, c \in (0, 1), \rho \in (0, 1), \bar{\alpha} > 0, \mu \geq 0$

for $k = 0, 1, \dots$ **do**

if $\|\tilde{\nabla} f(x^k)\| \geq \epsilon$ **then**

Set $B_k := \tilde{\nabla}^2 f(x^k)$ if $\tilde{\nabla}^2 f(x^k)_{-j, -j}$ is sufficiently p.d.

(delete rows/cols j if $x_j^k = 0$); otherwise $B_k := \mu \tilde{I}_k + \tilde{\nabla}^2 f(x^k)$.

Newton step: choose $d^k \in \mathbb{R}^n$ s.t. $-\tilde{\nabla} f(x^k) = B_k d^k$.

(continued)

Globalized Coderivative-Based Newtonian Algorithms for Single-Objective ℓ_0 Optimization

Algorithm 6 (continued): Armijo Backtracking Line Search Newton with Modification for ℓ_0 -Norm Optimization

Armijo backtracking: set $\alpha \leftarrow \bar{\alpha}$.
while $f(x^k + \alpha d^k) > f(x^k) + c\alpha \tilde{\nabla} f(x^k)^\top d^k$ **do**
 $\alpha \leftarrow \rho\alpha$.
end while
 Set $\alpha_k := \alpha$, $x^{k+1} := x^k + \alpha_k d^k$.
else stop.
end if
end for

Moreau Envelopes and Prox-Regularity

To deal with prox-regular (in fact strongly amenable) compositions in multiobjective optimization, denote the Moreau envelope restricted to hyperplane

$$e_r^{\tilde{I}(x)}\phi(x) := \inf_{w \in \tilde{I}(x)} \left\{ \phi(w) + \frac{1}{2r} \|w - x\|^2 \right\} = e_r(\phi + \delta_{\tilde{I}(x)})(x)$$

and define the set of directions

$$\Upsilon_{\tilde{I}(x), e_r\phi}^* := \{y \in \tilde{I}(x) \mid -\nabla e_r(\phi + \delta_{\tilde{I}(x)})(x) \in (D^*\nabla e_r(\phi + \delta_{\tilde{I}(x)}))(x)(-y)\}$$

Coderivative-Based Newtonian Algorithms for multiobjective ℓ_0 Optimization with Prox-Regular Compositions

Algorithm 7 coderivative-based Newtonian method for prox-regular compositions with ℓ_0 -norm

Require: Given x_0 as starting point, $\epsilon > 0$

for $k = 0, 1, \dots$

if $\|\tilde{\nabla} f(x^k)\| \geq \epsilon$ **then**

Newton step: Choose $d^k \in \mathbb{R}^n$ satisfying

$$d^k \in \Upsilon_{\tilde{I}(x), e_r \phi}^*(x^k)$$

Set $x^{k+1} := x^k - d^k$

else Stop

end if

Convergence Analysis and Numerical Experiment

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

Theorem 6.1.

*If we run **Algorithm 1**, there exists k -th iteration which is the iteration such that I_k becomes the smallest subspace, then it leads us to a local minimizer $\bar{x}$ satisfying the estimate*

$$f(x_{(k+s)}) + \|x_{(k+s)}\| - f(\bar{x}) - \|\bar{x}\| \leq \frac{\|x_k - \bar{x}\|_2^2}{2st}$$

Theorem 6.2.

Algorithm 3 converges a local minimizer $\bar{x}$ of **Weight sum**, which satisfies the estimate

$$\text{Weight sum}(x_{(k+s)}) + \|x_{(k+s)}\| - \text{Weight sum}(\bar{x}) - \|\bar{x}\| \leq \frac{\|x_k - \bar{x}\|_2^2}{2st}$$

Theorem 6.3.

In **Algorithm 4**, we have the estimate

$$\phi_A \circ F_{\text{best}}^{k+s} - \phi_A \circ F(\bar{x}) \leq \frac{\text{dist}(x^{(k+1)}, \bar{x})^2 + M^2 t^2 s}{2ts}$$

where M is the Lipschitz modulus of $\phi_A \circ F$

Theorem 6.4 (local quadratic convergence).

Consider the iteration in **Algorithm 5**, at some iteration k , $\tilde{I}(x^k)$ is the smallest subspace. Suppose that the Hessian $\tilde{\nabla}^2 f$ is Lipschitz continuous in a neighborhood of a solution x^* , which satisfies the second-order sufficient conditions, in the meaning of being embedded into subspace $\tilde{I}(x_k)$. Then

- i. If the point x^k is sufficiently close to x^* , the sequence of iterates converges to x^*
- ii. The rate of convergence of $\{x^{k+s}\}$ is *quadratic*
- iii. The sequence of gradient norms $\{\|\tilde{\nabla} f_{k+s}\|\}$ converges *quadratically* to zero

Theorem 6.5 (global convergence).

Let f be twice continuously differentiable on an open set $\mathcal{D}$, and assume that the starting point x_0 in **Algorithm 6** is such that the level set $\mathcal{L} = \{x \in \mathcal{D} : f(x) \leq f(x_0)\}$ is compact. Then we have

$$\lim_{k \rightarrow \infty} \tilde{\nabla} f(x^k) = 0$$

Suppose that the sequence of iterates x^k converges to a point x^* , where $\tilde{\nabla}^2 f(x^*)_{-j,-j}$ is positive-definite for sufficiently large k . We have that $\alpha_k = 1$ for all sufficiently large k , Algorithm 6 reduces to **Algorithm 5**, and the rate of convergence is *quadratic*.

Theorem 6.6 (solvability and superlinear convergence).

Let $\phi_A \circ F : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be *continuously prox-regular* at $\bar{x}$ for $\bar{v} = 0$, where $\bar{x}$ is a *tilt-stable local minimizer* for $\phi_A \circ F$ in the smallest subspace $\tilde{I}(x^k)$ over iterations with modulus $\kappa > 0$. Assume further that the mapping $\partial(\phi_A \circ F)$ is *semismooth** at $(\bar{x}, \bar{v})$. Then given any $r > 0$ sufficiently small, there exist $\delta > 0$ such that for $x^k \in \mathbb{B}_\delta(\bar{x}) \cap \tilde{I}(x^k)$ constructed by **Algorithm 7** converges to $\bar{x}$ and the rate of convergence is *superlinear*

Numerical Illustrations

- 1 Proton Arc Therapy
- 2 Mathematical Models
 - Multiobjective ℓ_0 -Norm Optimization
 - Single-Objective ℓ_0 -Norm Optimization
- 3 Preliminaries
 - Variational Analysis
 - Gerstewitz Scalarization
- 4 First-Order and Second-Order Algorithms
 - Subgradient-Based Algorithms for Single-Objective ℓ_0 Optimization
 - Subgradient-Based Algorithms for Multiobjective ℓ_0 Optimization
- 5 Coderivative-Based Newtonian Algorithms
 - Algorithm for Single-Objective ℓ_0 Optimization
 - Algorithms for Multiobjective ℓ_0 -Norm Optimization
- 6 Convergence Analysis and Numerical Experiments
 - Convergence Analysis
- 7 Numerical Illustrations

Algorithm 1

EXAMPLE

Let $f(x, y) := x^2 + 2y^2 - 2x - 2xy + 3 + \|(x, y)\|_0$ for $(x, y) \in \mathbb{R}^2$.

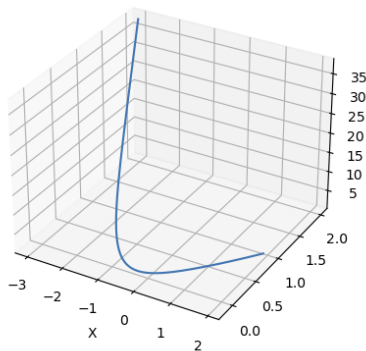


Figure: Initial point $(-3, 2)$

Algorithm 1

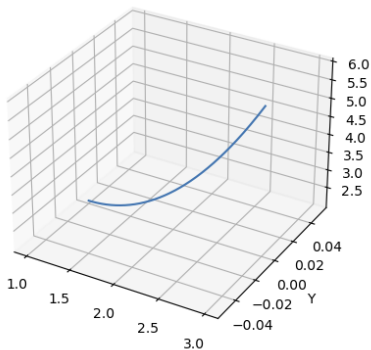


Figure: Initial point $(3,0)$

Algorithm 1

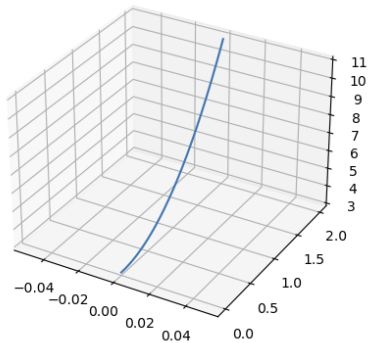


Figure: Initial point $(0, 2)$

Algorithm 1 with alternative definition for ℓ_0

Instead of defining the ℓ_0 -norm by the condition that x_i exactly equal to zero, we modify it by $|x_i| \leq \epsilon$, with $\epsilon = 10^{-6}$ in the next figure, to see the differences

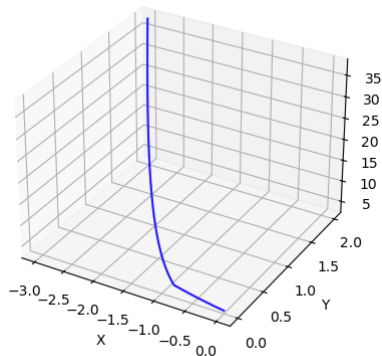


Figure: Initial point $(-3, 2)$ with alternative definition

Algorithm 1 with alternative definition for ℓ_0

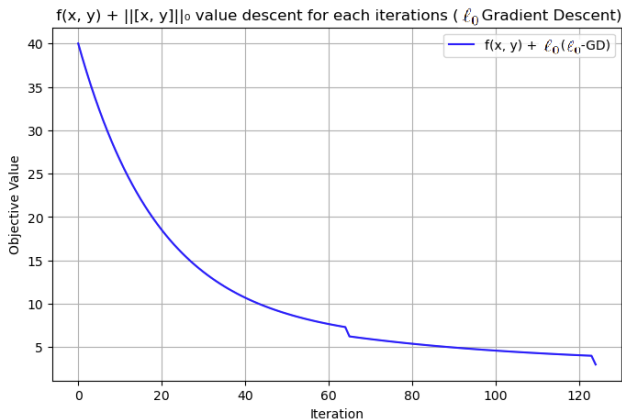


Figure: Adaptive definition for the ℓ_0 gradient descent

EXAMPLE Let $f_1(x, y) := (x - 1)^2 + y^2$ and $f_2(x) := x^2 + (y - 2)^2$. We choose the set $A = \{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 + y_2 \leq 0\}$ with $k^0 = (1, 1)$ in Algorithm 4. The definition of the Gerstewitz function gives us $\phi_A(f_1, f_2) = \inf\{t \mid (f_1, f_2) - tk^0 \in A\}$, and so $f_1 - t + f_2 - t \leq 0$, which yields $t \geq \frac{f_1 + f_2}{2}$. Taking the infimum of t , we get $t = \frac{f_1 + f_2}{2}$. Therefore, this scalarization agrees with the case of the weighted-sum method

Algorithm 3

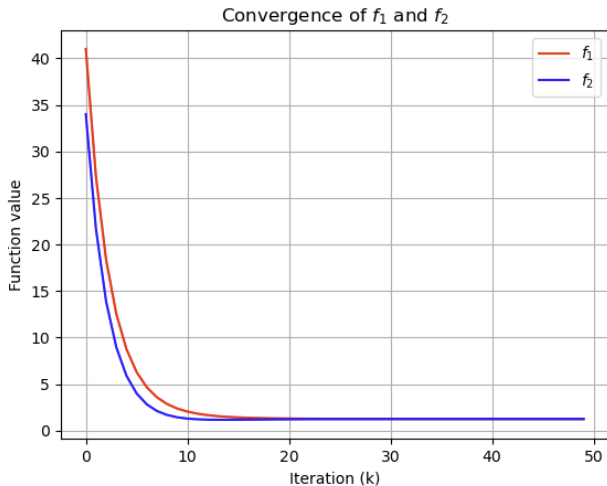


Figure: Weighted-Sum Method

EXAMPLE. Let $f_1(x) = (x - 2)^2$ and $f_2(x) := (x + 1)^2 + 1$. In this example, we choose the set $A := \{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 \text{ and } y_2 \leq 0\}$. We have $\phi_A(f_1, f_2) = \inf\{t \mid (f_1, f_2) - tk^0 \in A\}$. Then $t \geq f_1$ and $t \geq f_2$. Thus $t \geq \max\{f_1, f_2\}$ with the infimum of t calculated by $t = \max\{f_1, f_2\}$.

Algorithm 4

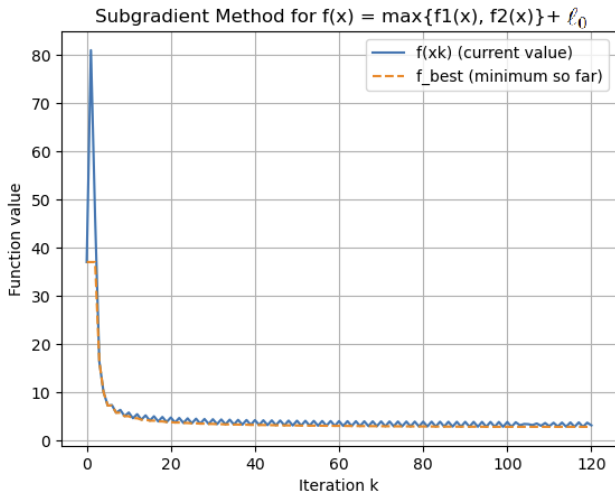


Figure: Maximum Function

Algorithm 6

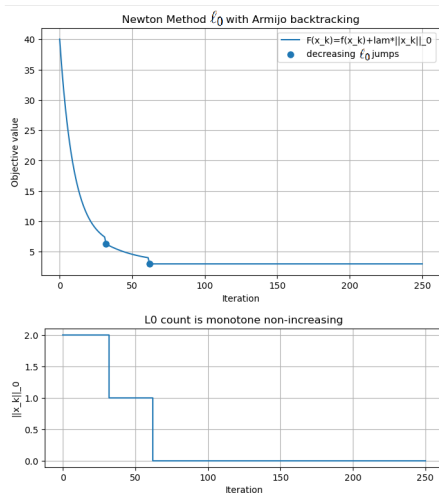


Figure: Armijo backtracking Newton method

Algorithm 6 and Algorithm 1 comparison

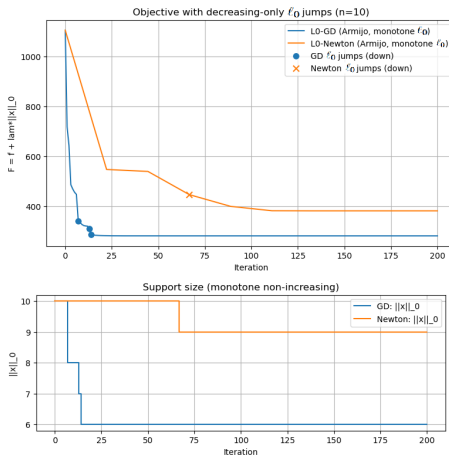


Figure: Newton method and Gradient descent Comparison

Algorithm 7

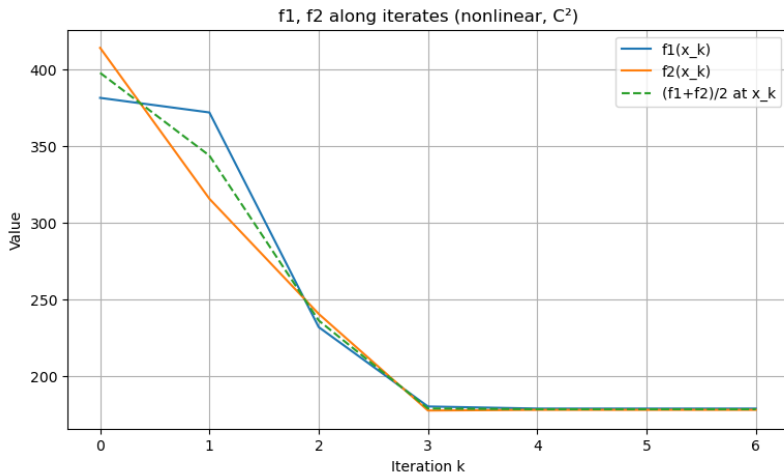


Figure: Weight sum Newton method

Algorithm 7

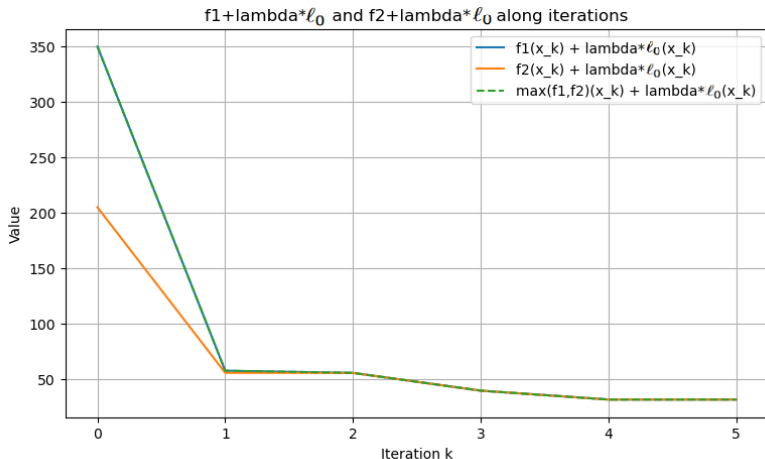


Figure: Newton method for maximum function

More detail on the [first-order algorithms](#) can be found in the paper

X. Cong, X. Ding, B. S. Mordukhovich, A. Nguyen and L. Zhao, ℓ_0 -Norm multiobjective optimization models motivated by applications to proton therapy, *J. Nonlin. Var. Anal.* 6 (2026), arXiv:2509.06833

The [second-order algorithms](#) are given in

B. S. Mordukhovich and A. Nguyen, [Coderivative-based algorithms for sparsity optimization](#), preprint (2026)

All of this is presented in the new book

B. S. Mordukhovich, [Fundamentals of Variational Analysis](#), to appear in Springer (2027)