Andy's Early Work: 1971 – 1982

ICCOPT Berlin 2019

Doctoral Thesis: University of Waterloo 1971

A Gradient Type Method for Locating Constrained Extrema

Advisor: Tomasz Pietrzykowski

Research Area: Exact Penalization in Nonlinear Programming

The extension of steepest descent to nonsmooth optimization and the origins of *vertical* and *horizontal* steps.

Exact Penalization

$$
\begin{array}{llrl}
\text{NLP} & \text{minimize} & f(x) & \\
& \text{subject to} & \phi_i(x) \leq & 0 \quad i = 1, \ldots, k \\
& & \phi_i(x) = & 0 \quad i = k+1, \ldots, \ell
\end{array}
$$

where f and all ϕ_i are continuous mappings from $\mathbb{R}^n$ to $\mathbb{R}$.

Feasible region:

$$
\mathcal{F} := \left\{ x \;\middle|\; \begin{array}{r} \phi_i(x) \leq 0,\ i = 1, \ldots, k, \\ \phi_i(x) = 0,\ i = k+1, \ldots, \ell \end{array} \right\}
$$

Exact Penalization

$$\begin{array}{llrl} \text{NLP} & \text{minimize} & f(x) & \\ & \text{subject to} & \phi_i(x) \leq & 0 \quad i = 1, \ldots, k \\ & & \phi_i(x) = & 0 \quad i = k+1, \ldots, \ell \end{array}$$

where f and all ϕ_i are continuous mappings from $\mathbb{R}^n$ to $\mathbb{R}$.

Feasible region:

$$\mathcal{F} := \left\{ x \;\middle|\; \begin{array}{r} \phi_i(x) \leq 0,\ i = 1, \ldots, k, \\ \phi_i(x) = 0,\ i = k+1, \ldots, \ell \end{array} \right\}$$

ℓ_1 Exact Penalization:

$$\ell_1\text{–NLP} \qquad \min p_\mu(x) := \mu f(x) + \sum_{i=1}^{k} \max(0, \phi_i(x)) + \sum_{i=k+1}^{\ell} |\phi_i(x)|$$

Theoretical Foundations

Thm:(CQ) If $\bar{x}$ solves NLP, then, for all $\mu > 0$ small, $\bar{x}$ solves ℓ_1-NLP:

$$\ell_1 - \text{NLP} \quad \min_x \, p_\mu(x) := \mu f(x) + \sum_{i=1}^{k} \max(0, \phi_i(x)) + \sum_{i=k+1}^{\ell} |\phi_i(x)| \; .$$

Theoretical Foundations

Thm:(CQ) If $\bar{x}$ solves NLP, then, for all $\mu > 0$ small, $\bar{x}$ solves ℓ_1-NLP:

$$\ell_1 - \text{NLP} \quad \min_x \, p_\mu(x) := \mu f(x) + \sum_{i=1}^{k} \max(0, \phi_i(x)) + \sum_{i=k+1}^{\ell} |\phi_i(x)| \, .$$

Convex Case (finite-valued): Eremin (1966), Zangwill (1967)

Slater CQ: ϕ_i are affine for $i = k+1, \ldots, \ell$ and
$\exists \, \hat{x} \in \mathcal{F}$ such that $\phi_i(\hat{x}) < 0, \; i = 1, \ldots, k$.

Theoretical Foundations

Thm:(CQ) If $\bar{x}$ solves NLP, then, for all $\mu > 0$ small, $\bar{x}$ solves ℓ_1-NLP:

$$\ell_1 - \text{NLP} \quad \min_x \; p_\mu(x) := \mu f(x) + \textstyle\sum_{i=1}^{k} \max(0, \phi_i(x)) + \sum_{i=k+1}^{\ell} |\phi_i(x)| \; .$$

Convex Case (finite-valued): Eremin (1966), Zangwill (1967)

Slater CQ: ϕ_i are affine for $i = k+1, \ldots, \ell$ and
$\exists \; \hat{x} \in \mathcal{F}$ such that $\phi_i(\hat{x}) < 0, \;\; i = 1, \ldots, k.$

Smooth Case: Pietrzykowski (1969)

(LICQ): The active constraint gradients,

$$\nabla \phi_i(x) \quad i \in A(x, 0), \qquad \text{are } \textit{linearly independent},$$

where, for $\varepsilon \geq 0$,

$$A(x, \varepsilon) := \{ i \mid |\phi_i(x)| \leq \varepsilon \, , \;\; i \in \{1, \ldots, k\} \}$$

are the ε-active constraints.

Vertical and Horizontal Steps

Constrained Optimization Using a Nondifferentiable Penalty Function,
SIAM J. Numerical Analysis, 10(1973)760–784.

Linear Programming via a Nondifferentiable Penalty Function
SIAM J. Numerical Analysis, 13(1976)145–154.

A Penalty Function Method Converging Directly to a Constrained Optimum
with Tomasz Pietrzykowski
SIAM J. Numerical Analysis, 14(1977)348–375.

Vertical and Horizontal Steps

For simplicity assume $\mathcal{F} := \{x \mid \phi_i(x) \le 0,\ i = 1, \ldots, \ell\}$.

$$A(x,\varepsilon) := \{i \mid |\phi_i(x)| \le \varepsilon\,,\ i \in \{1,\ldots,\ell\}\} \qquad \varepsilon\text{-active}$$

$$I(x,\varepsilon) := \{i \mid |\phi_i(x)| > \varepsilon\,,\ i \in \{1,\ldots,\ell\}\} \qquad \varepsilon\text{-inactive}$$

$$\widehat{I}(x,\varepsilon) := I(x,\varepsilon) \cap \{i \mid \phi_i(x) > 0,\ i = 1,\ldots,\ell\} \qquad \text{infeas. } \varepsilon\text{-inactive}$$

Keys: The construction of P and the evaluation of $\sigma,\ \tau \ge 0$.

Vertical and Horizontal Steps

For simplicity assume $\mathcal{F} := \{x \mid \phi_i(x) \le 0,\ i = 1, \dots, \ell\}$.

$$A(x,\varepsilon) := \{i \mid |\phi_i(x)| \le \varepsilon\,,\ i \in \{1,\dots,\ell\}\} \qquad \varepsilon\text{-active}$$
$$I(x,\varepsilon) := \{i \mid |\phi_i(x)| > \varepsilon\,,\ i \in \{1,\dots,\ell\}\} \qquad \varepsilon\text{-inactive}$$
$$\widehat{I}(x,\varepsilon) := I(x,\varepsilon) \cap \{i \mid \phi_i(x) > 0,\ i = 1,\dots,\ell\} \qquad \text{infeas. } \varepsilon\text{-inactive}$$

"Steepest Descent" for p_μ: $r(x,\varepsilon) := -\mu\nabla f(x)\ - \sum_{i\in\widehat{I}(x,\varepsilon)} \nabla\phi_i(x)$

Keys: The construction of P and the evaluation of $\sigma,\ \tau\ \ge 0$.

Vertical and Horizontal Steps

For simplicity assume $\mathcal{F} := \{x \mid \phi_i(x) \le 0,\ i = 1, \ldots, \ell\}$.

$$A(x,\varepsilon) := \{i \mid |\phi_i(x)| \le \varepsilon\,,\ i \in \{1,\ldots,\ell\}\} \qquad \varepsilon\text{-active}$$
$$I(x,\varepsilon) := \{i \mid |\phi_i(x)| > \varepsilon\,,\ i \in \{1,\ldots,\ell\}\} \qquad \varepsilon\text{-inactive}$$
$$\widehat{I}(x,\varepsilon) := I(x,\varepsilon) \cap \{i \mid \phi_i(x) > 0,\ i = 1,\ldots,\ell\} \qquad \text{infeas. } \varepsilon\text{-inactive}$$

"Steepest Descent" for p_μ: $r(x,\varepsilon) := -\mu\nabla f(x) \ - \sum_{i\in\widehat{I}(x,\varepsilon)} \nabla\phi_i(x)$

Let P be (almost) the projection onto the subspace orthogonal to the ε-active constraint gradients:

$$\mathrm{Span}[\{\nabla\phi_i(x) \mid i \in A(x,\varepsilon)\}]^\perp.$$

Keys: The construction of P and the evaluation of $\sigma,\ \tau \ge 0$.

Vertical and Horizontal Steps

For simplicity assume $\mathcal{F} := \{x \mid \phi_i(x) \leq 0,\ i = 1, \ldots, \ell\}$.

$$
\begin{array}{ll}
A(x,\varepsilon) := \{i \mid |\phi_i(x)| \leq \varepsilon\,,\ i \in \{1, \ldots, \ell\}\} & \varepsilon\text{-active} \\
I(x,\varepsilon) := \{i \mid |\phi_i(x)| > \varepsilon\,,\ i \in \{1, \ldots, \ell\}\} & \varepsilon\text{-inactive} \\
\widehat{I}(x,\varepsilon) := I(x,\varepsilon) \cap \{i \mid \phi_i(x) > 0,\ i = 1, \ldots, \ell\} & \text{infeas. } \varepsilon\text{-inactive}
\end{array}
$$

"Steepest Descent" for p_μ: $r(x,\varepsilon) := -\mu \nabla f(x) \ - \sum_{i \in \widehat{I}(x,\varepsilon)} \nabla \phi_i(x)$

Let P be (almost) the projection onto the subspace orthogonal to the ε-active constraint gradients:

$$\mathrm{Span}[\{\nabla \phi_i(x) \mid i \in A(x,\varepsilon)\}]^\perp .$$

$$
\begin{array}{ll}
h(x,\varepsilon) := P\, r(x,\varepsilon) & \text{the horizontal step} \\
v(x,\varepsilon) := (I - P) r(x,\varepsilon) & \text{the vertical step} \\
w(x,\varepsilon) := \sigma\, v(x,\varepsilon) \ + \ \tau\, v(x,\varepsilon) & \text{the step}
\end{array}
$$

Keys: The construction of P and the evaluation of $\sigma,\ \tau \geq 0$.

Extensions

> *UV-decompositions are an example of recent ideas in this direction, where the horizontal step is in the U direction and the vertical step is in the V direction.*

Minimization Techniques for Piecewise Differentiable Functions: The ℓ_1 Solution to an Overdetermined Linear System
with Richard Bartels and James Sinclair
SIAM J. Numerical Analysis, 15(1978)224–241.

Linearly Constrained Discrete ℓ_1 Problems
with Richard Bartels
AMS TOMS 4(1980)594–608.

An Efficient Method to Solve the MiniMax Problem Directly
with Christakas Charalambous
SIAM J. Numerical Analysis, 15(1978)162–241.

Second-Order Theory and Algorithms

Second-Order Conditions for and Exact Penalty Function
with Tom Coleman
Mathematical Programming 19(1980)178–185.

Nonlinear Programming via and Exact Penalty Function: Asymptotic Analysis
with Tom Coleman
Mathematical Programming 24(1982)123–136.

Nonlinear Programming via and Exact Penalty Function: Global Analysis
with Tom Coleman
Mathematical Programming 24(1982)137–161.

Second-Order Theory and Algorithms

Theory:
Andy and Tom established second-order necessary and sufficient conditions for the ℓ_1 exact penalty function using techniques from NLP under LICQ.

- The theory applies at both feasible and infeasible points.
- When feasible, they show equivalence with the NLP strong second-order theory.

Second-Order Theory and Algorithms

Algorithms:

Again, the basic idea rests on the notion of vertical and horizontal steps.

But now the horizontal step h^k is based on a second-order approximation to the Lagrangian over the subspace perpendicular to the active constraint gradients.

Multiplier estimates are given by a least-squares solution to the first-order optimality conditions.

Second-Order Theory and Algorithms

Algorithms:

Again, the basic idea rests on the notion of vertical and horizontal steps.

But now the horizontal step h^k is based on a second-order approximation to the Lagrangian over the subspace perpendicular to the active constraint gradients.

Multiplier estimates are given by a least-squares solution to the first-order optimality conditions.

Once the second-order step is chosen, a vertical step v^k is chosen at the point $x^k + h^k$ using the data at x^k to give the final step $x^k + h^k + v^k$.

Second-Order Theory and Algorithms

Algorithms:

Again, the basic idea rests on the notion of vertical and horizontal steps.

But now the horizontal step h^k is based on a second-order approximation to the Lagrangian over the subspace perpendicular to the active constraint gradients.

Multiplier estimates are given by a least-squares solution to the first-order optimality conditions.

Once the second-order step is chosen, a vertical step v^k is chosen at the point $x^k + h^k$ using the data at x^k to give the final step $x^k + h^k + v^k$.

This work is one of the initial contributions toward second-order correction steps (Fletcher) to overcome the Marotos effect.

Second-Order Theory and Algorithms

Convergence Theory:

Local: Andy and Tom establish the two step local super-linear convergence of their method under a strong second-order sufficiency.

Global:

- A break-point line-search procedure is introduced to ensure global convergence.
- Under a strong second-order sufficiency condition, the *Newton* step is accepted and two step super-linear convergence is achieved.

Thank You Andy!!

An inspiring leader, mentor, community builder, and researcher.