

Problemstellung und Motivation

Die l_1 -Minimierung ist ein noch recht junges Teilgebiet der Nichtlinearen Optimierung, das Probleme der Form

$$(1) \quad \min_{x \in \mathbb{R}^n} \mu \cdot \|x\|_1 + f(x)$$

für eine konvexe und in der Regel stetig differenzierbare Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ und $\mu > 0$ betrachtet. Neue Ergebnisse in der Signalverarbeitung motivieren solche nichtglatten Probleme und beruhen auf der Erkenntnis, dass die charakterisierenden und markanten Informationen eines Signals oder Bildes im Vergleich zum gesamten Signal oder Bild eine sehr dünn besetzte Darstellung besitzen. Ist der Vektor $\bar{x} \in \mathbb{R}^n$ ein dünn besetztes Signal, so kann es aus einer verhältnismäßig geringen Anzahl an Messungen $b = A\bar{x}$ unter einer geeigneten Wahl von $A \in \mathbb{R}^{m \times n}$, $m \leq n$, mit Hilfe des l_1 -Optimierungsproblems

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{u.d.N. } Ax = b$$

rekonstruiert werden. Dieses Minimierungsproblem kann leicht in die Form (1) gebracht werden. Wir stellen nun für die obige, allgemeine Klasse von Problemen ein neues, schnell lokal konvergentes Verfahren vor, das auf einem semiglattem Newton-Verfahren basiert und das mit Hilfe eines abstrakten Filter-Konzeptes globalisiert wird.

Optimalitätsbedingungen

Fixpunkt-Gleichung. Sei X^* die Menge der optimalen Lösungen von (1) und $\tau > 0$ beliebig, dann gilt

$$(2) \quad x \in X^* \Leftrightarrow x = \text{sgn}(x - \tau g(x)) \odot \max\{|x - \tau g(x)| - \tau\mu, 0\}.$$

Die Fixpunkt-Gleichung (2) kann als Verknüpfung der beiden Abbildungen

- $h(\cdot) = I(\cdot) - \tau g(\cdot)$
- $\mathcal{S}_{\tau\mu}(\cdot) = \text{sgn}(\cdot) \odot \max\{|\cdot| - \tau\mu, 0\}$

geschrieben werden, wobei $g = \nabla f$ den Gradienten von f und „ \odot “ ein komponentenweises Vektorprodukt bezeichnet. $\mathcal{S}_{\tau\mu}$ wird wegen seiner einschrumpfenden Eigenschaft auch **Shrinkage-Operator** genannt. Aus der Optimalitätsbedingung (2) ergibt sich folgende Fixpunktiteration

$$x^{k+1} = \mathcal{S}_{\tau\mu} \circ h(x^k) \quad \text{mit } \tau > 0.$$

Letztere Vorschrift wurde intensiv in [2] untersucht und kann zu einem global konvergenten Abstiegsverfahren mit $x^{k+1} = x^k + \sigma_k d^k$ erweitert werden. Hierbei stellt $d^k = \mathcal{S}_{\tau\mu} \circ h(x^k) - x^k$ eine Abstiegsrichtung und σ_k eine geeignete Schrittweite dar.

Nullstellen-Gleichung. Die Optimalitätsbedingung (2) ist äquivalent zu

$$(3) \quad x \in X^* \Leftrightarrow F_\tau(x) = 0,$$

wobei $F_\tau(x) := g(x) - \mathcal{P}_{[-\mu, \mu]^n}(g(x) - \frac{x}{\tau})$ gilt.

Ein mehrdimensionales Filter-Konzept

Für eine (nichtglatte) Funktion $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ betrachten wir das Nullstellenproblem

$$(4) \quad F(x) = 0$$

und wollen wie folgt einen **Filtereintrag** konstruieren (vergleiche [1]):

- Zerlege (4) in p möglicherweise überlappende Gleichungen $(F_{I_j}(x))$ für $j \in \{1, \dots, p\}$ mit $\{1, 2, \dots, m\} = \bigcup_{j=1}^p I_j$.
- Definiere $\theta: \mathbb{R}^n \rightarrow \mathbb{R}_+^p$ durch $\theta_j(x) := \|F_{I_j}(x)\|_2$ für alle $j \in \{1, \dots, p\}$.
- Beispiele: $p = 1$, $\theta = \|F\|$ oder $p = m$, $\theta = (|F_1|, \dots, |F_m|)^T$.

Ein Filter \mathcal{F} ist nun eine spezielle **Speicherstruktur**, d.h. eine Liste aus endlich vielen Vektoren $\mathcal{F} := \{\theta_{\mathcal{F}}^1, \theta_{\mathcal{F}}^2, \dots, \theta_{\mathcal{F}}^p\}$ mit $\theta_{\mathcal{F}}^j \in \mathbb{R}_+^p$.

Filter-Zulässigkeit. Ein Vektor $\vartheta \in \mathbb{R}_+^p$ heißt **zulässig für den Filter \mathcal{F}** , falls

$$\max_{1 \leq j \leq p} \theta_{\mathcal{F},j} - \vartheta_j \geq \gamma_{\mathcal{F}} \cdot \|\vartheta\|_\infty \quad \text{für alle } \theta_{\mathcal{F}} \in \mathcal{F}$$

gilt, wobei $\gamma_{\mathcal{F}}$ eine Konstante aus dem Intervall $(0, 1)$ ist.

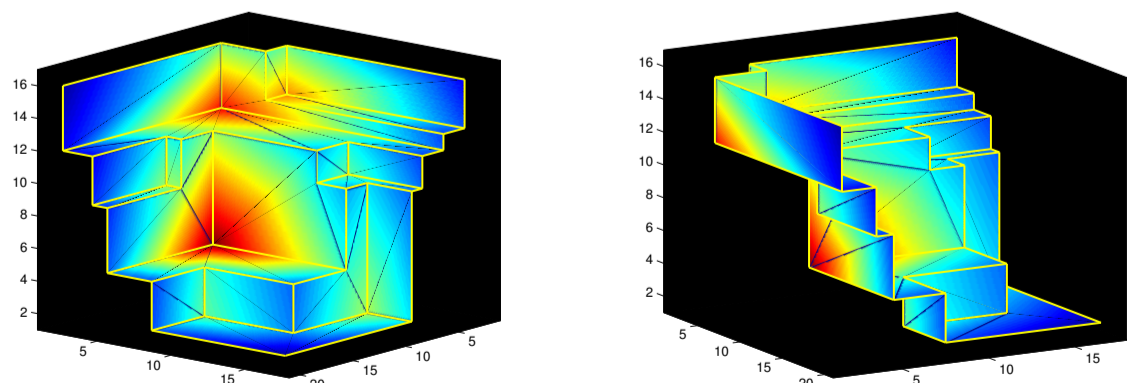


Abbildung 1: Beispiel eines dreidimensionalen Filters im Grenzfall $\gamma_{\mathcal{F}} = 0$. Alle Vektoren, die hinter der farbigen Grenzwannd (also näher an der Null) liegen, sind zulässig für den Filter.

Die Idee des Filterverfahrens besteht also darin, die m verschiedenen Komponenten F_i unabhängig voneinander bzw. in den einzelnen Gruppen I_j gegen Null zu schicken. Durch die Zulässigkeitsbedingung wird nur dann ein Vektor in den Filter aufgenommen, wenn er in mindestens einer (unterschiedlichen) Komponente besser als die vorangegangenen Einträge ist. Wir nähern uns also „gruppenweise“ der Null. Dieses abstrakte Filter-Konzept kann zur Globalisierung von Algorithmen genutzt werden.

Semiglatte Newton-Verfahren

Wir betrachten wieder das Problem (4) und bezeichnen mit D_F , die Menge der Punkte, an denen F differenzierbar ist. Ist die Funktion F lokal Lipschitz-stetig, so folgt aus dem **Satz von Rademacher**, dass F fast überall differenzierbar ist.

Verallgemeinerte Differentiale. Sei $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz-stetig in einer Umgebung von x . Die Menge

$$\partial_B F(x) := \{M \in \mathbb{R}^{m \times n} : \exists (x^k) \subset D_F \text{ mit } x^k \rightarrow x, F'(x^k) \rightarrow M\}$$

heißt **Bouligand-Subdifferential** oder **B-Subdifferential** von F in x . Das **Clarke-Subdifferential** von F in x ist die konvexe Hülle $\partial F(x) := \text{co}(\partial_B F(x))$.

Semiglattheit. Sei $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz-stetig in einer Umgebung von x . Dann heißt F **semiglat** in x , wenn $F'(x, h)$ für alle $h \in \mathbb{R}^n$ existiert und

$$\sup_{M \in \partial F(x+h)} \|F(x+h) - F(x) - Mh\| = o(\|h\|).$$

Das semiglatte Newton-Verfahren basiert auf den letzten Definitionen und berechnet ausgehend von x^k eine neue iterierte $x^{k+1} = x^k + s^k$ mit $M_k s^k = -F'(x^k)$. Da wir die Funktion F nicht als glatt voraussetzen wollen, greifen wir auf die Subdifferentialia $M_k \in \partial F(x^k)$ zurück. Stellt x^* nun eine Nullstelle von F dar, ist F semiglat in x^* und sind alle Matrizen $M \in \partial F(x^*)$ nichtsingulär, dann konvergiert die Folge (x^k) q -superlinear gegen x^* , falls der Startpunkt x^0 nahe an x^* liegt (siehe [4]). Da die Funktion F_τ als stückweise stetig differenzierbare Funktion semiglat ist, können wir das semiglatte Newton-Verfahren zur Lösung von (3) verwenden. Mit Hilfe des Filter-Konzeptes ergibt sich folgende, globalisierte Variante des Newton-Verfahrens.

Semiglatte Newton-Verfahren mit mehrdim. Filter-Globalisierung.

0. Initialisierung: Wähle $x^0 \in \mathbb{R}^n$, $\tau > 0$, $\gamma_{\mathcal{F}} \in (0, 1)$, $\mathcal{F} = \emptyset$ und $k = 0$
1. Berechne eine Newton-Richtung $M(x^k)s^k = -F_\tau'(x^k)$ (falls nicht möglich, lasse Schritt 3 aus)
2. Falls $F_\tau(x^k) = 0$, STOP
3. Setze $x^{k+1} = x^k + s^k$ und überprüfe, ob $|F_\tau(x^{k+1})|$ für den Filter zulässig ist. Falls $|F_\tau(x^{k+1})|$ zulässig ist, aktualisiere \mathcal{F} , erhöhe k und gehe zu Schritt 1
4. Berechne $x^{k+1} = x^k + \sigma_k d^k$ über das globalisierte Fixpunktverfahren, erhöhe k und gehe zu Schritt 1

Sei (x^k) vom obigen Algorithmus erzeugt, dann erhalten wir (vergleiche [3]):

Globale Konvergenz I. Ist x^* ein Häufungspunkt der Folge (x^k) und existiert eine Teilfolge $(x^k)_K$, die gegen x^* konvergiert und unendlich viele Newton-Schritte verwendet, dann ist x^* ein globales Minimum.

Globale Konvergenz II. Die Lösungsmenge X^* sei nichtleer, kompakt und $\Omega \supset X^*$ sei hinreichend groß. Für die konvexe Funktion f gelte $f \in C^2(\Omega)$ und die Hessematrix H von f sei auf Ω beschränkt, dann gilt für einen beliebigen Startpunkt:

- Die Folge (x^k) ist beschränkt.
- Jeder Häufungspunkt von (x^k) ist ein globales Minimum.

Schnelle lokale Konvergenz. Zusätzlich zu den Annahmen im letzten Satz sei die quadratische Untermatrix $H_{\mathcal{A}_* \mathcal{A}_*}(x^*)$ für jeden Häufungspunkt $x^* \in X^*$ der Folge (x^k) und für $\mathcal{A}_* := \{i : |g_i(x^*)| = \mu\}$ positiv definit, dann gilt:

- Die Folge (x^k) konvergiert gegen ein eindeutiges Minimum der Zielfunktion.
- Das Verfahren geht in das semiglatte Newton-Verfahren über.

Literatur

- [1] Nicholas I. M. Gould, Sven Leyffer und Philippe L. Toint, *A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares*, SIAM Journal on Optimization, 15 (2004), Seiten 17-38.
- [2] Elaine T. Hale, Wotao Yin und Yin Zhang, *Fixed-point continuation for l_1 -minimization: methodology and convergence*, SIAM Journal on Optimization, 19 (2008), Seiten 1107-1130.
- [3] Andre Milzarek, *Ein semiglatte Newton-Verfahren mit mehrdimensionaler Filter-Globalisierung zur Lösung von l_1 -Minimierungsproblemen*, Bachelorarbeit, TU München 2010.
- [4] Liqun Qi und Jie Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), Seiten 353-367.