Christian Clason. Foto: Vladimir Unkovic

Nichtglatte Optimierung beschäftigt sich mit der Minimierung von Funktionen, die nicht im klassischen Sinn differenzierbar sind. Christian Clason gibt einen Einblick in moderne Verfahren und ihre Anwendung in der mathematischen Bildverarbeitung.

Optimierung mit „Knick“

Nichtglatte Optimierung in der Bildverarbeitung

Von Christian Clason

Viele Probleme aus Naturwissenschaft, Technik, und Wirtschaft lassen sich mathematisch als Optimierungsprobleme formulieren: Gesucht wird ein Objekt, für das eine vorgegebene *Kostenfunktion* den minimalen (oder, äquivalent durch Betrachtung der negativen Kostenfunktion, maximalen) Wert annimmt. In der Physik sind zum Beispiel zahlreiche Beobachtungen dadurch erklärt, dass eine geeignete Energie minimiert wird. In der Wirtschaft ist der Zusammenhang noch direkter: Produktionspläne sollen die tatsächlich anfallenden Produktions- und Transportkosten minus Gewinn minimieren (daher auch der Name), und Portfolios sollen so zusammengestellt werden, dass der Verlust über die Laufzeit minimiert wird. Um den gesuchten Minimierer zu finden, geht man in der Regel analog zu der aus der Schule bekannten Kurvendiskussion vor: Man setzt die Ableitung gleich Null und löst auf; ist die zweite Ableitung in dem gefundenen Punkt positiv, hat man einen Minimierer. Eine der wesentlichen Leistungen der Mathematik zu Anfang des 20. Jahrhunderts war es zu ermöglichen, dass eine breite Klasse von Objekten – darunter Funktionen selber – als Punkte in einem geeigneten Vektorraum aufgefasst werden können, der die notwendigen Strukturen aufweist, um Funktionen darauf definieren und ableiten zu können. Dies ist heute als *Funktionalanalysis* und *Variationsrechnung* fester Bestandteil des Mathematikstudiums.

Was macht man aber, wenn die Kostenfunktion nicht differenzierbar ist, also keine Ableitung im klassischen Sinn existiert? Paradebeispiel ist hier die Betragsfunktion, die in ihrem „Knick" in der Null keine Ableitung besitzt. (Dieses Beispiel wird uns als roter Faden durch den gesamten Artikel begleiten.) Bevor wir dazu kommen, wenden wir uns aber erstmal der – durchaus zulässigen – Frage zu, *warum* man solche Funktionen überhaupt minimieren möchte.

Compressed Sensing

Tatsächlich hat die Minimierung von nichtdifferenzierbaren Funktionen eine lange Tradition. Spätestens seit der Mitte des letzten Jahrhunderts beschäftigen sich insbesondere die Wirtschaftswissenschaften mit der Optimierung von Funktionen, die „Knicke" aufweisen. Ein typisches Beispiel sind *stückweise* lineare Funktionen; man denke etwa an die Einkommenssteuer mit einem Steuersatz, der sich bei Überschreiten eines bestimmten Jahreseinkommens erhöht. Wesentliche neue Impulse hat dieses Thema aber in den letzten Jahren durch neue Anwendungen in der mathematischen Bildverarbeitung erhalten, und auf diese konzentrieren wir uns in Folge.

Ein zentraler Auslöser dafür war die Theorie des *compressed sensing*. Ausgangspunkt war hier die Frage, unter welchen Zusatzannahmen ein *unterbestimmtes* lineares Gleichungssystem – das im Allgemeinen unendlich viele Lösungen besitzt – eine (irgendwie) ausgezeichnete eindeutige Lösung bezeichnet. Die verblüffende Antwort ist, dass dies möglich ist, wenn man speziell eine Lösung sucht, die *sparse* (in etwa „dünn besetzt", „spärlich") ist, das heißt wenn nur wenige Komponenten des gesuchten Vektors von Null verschieden sind (und die Koeffizienten des Gleichungssystems bestimmte Eigenschaften haben).[1] Der Beweis dieser Aussage ist nicht einfach, hat aber elegante Verbindungen zur Theorie der Zufallsmatrizen. Noch überraschender – und für uns relevanter – ist die Tatsache, dass diese *sparse* Lösung diejenige unter allen Lösungen ist, die *minimale 1-Norm* (das heißt die Summe der Absolutbeträge der Komponenten) hat. Man kann diese Lösung also tatsächlich berechnen durch Lösen eines modifizierten Ausgleichsproblems, das aus der Summe des klassischen kleinste-Quadrate-Terms und der (geeignet gewichteten) 1-Norm besteht. Auch hier ist der Beweis technisch; die folgende anschauliche Erklärung soll das aber zumindest motivieren. Wir betrachten nur den Fall einer linearen Gleichung für eine Unbekannte; hier ist die Lösung sparse genau dann, wenn sie Null ist – also im Knick des Absolutbetrags (d. h. der 1-Norm in diesem Fall) liegt. Angenommen, wir könnten jetzt den Ausgleichsterm kleiner machen, indem wir zu der Null eine kleine Zahl addieren. Da der Ausgleichsterm quadratisch ist, ist er dann um das Quadrat dieser Zahl kleiner. Und da das Quadrat einer Zahl kleiner als Eins kleiner ist als die Zahl selber, ist die Gesamtänderung der zu minimierenden Summe aus Ausgleichsterm und 1-Norm positiv – wir finden also sicher keinen Minimierer, indem wir den Knick verlassen. Eine Lösung wird also im Allgemeinen immer in den Knicken einer stückweise linearen Funktion zu finden sein; Analoges gilt für die Minimierung der Summe von Euklidischer und 1-Norm. (Ein ähnliches Phänomen findet sich in der linearen Optimierung, wo eine Lösung immer in einer Ecke des zulässigen Bereichs liegt.)

Mathematische Bildverarbeitung

Diese Erkenntnis hatte auch Auswirkungen auf die *mathematische Bildverarbeitung*. Diese beschäftigt sich mit Verfahren, um aus verrauschten, unscharfen, und/oder unvollständigen Aufnahmen eine Näherung des ursprünglichen Bilds zu berechnen. Eng verwandt ist das Gebiet der *Bildgebung*, wo die gegebenen Daten nicht selber die Form von Bildern haben; klassische Beispiele sind die Computertomographie (CT) und die magnetische Resonanztomographie (MRT). Neben den Fourier- oder Wavelet-basierten Filterverfahren haben sich dort in den letzten Jahren sogenannte *Variationsmethoden* durchgesetzt. Dort fasst man Bilder als Funktionen auf, die einem Punkt (bzw. einem Pixel) den entsprechenden Farbwert (oder einfacher, Grauwert für Schwarz-Weiß-Bilder) zuordnet, und sucht das – in einem geeigneten Funktionenraum – *optimale* Bild, das am besten zu den gegebenen Daten passt. Dies führt auf die Minimierung der (gewichteten) Summe eines *Diskrepanzterms*, der die Nähe zu den gegebenen Daten misst, und eines *Regularisierungsterms*, der messen soll, wie „gut" das Bild an sich ist. Der Regularisierungsterm wird dabei vor allem durch die Wahl des Funktionenraums bestimmt. Die Gewichtung bestimmt, was uns dabei wichtiger ist: gute Übereinstimmung zu den Daten oder intrinsische „Güte" der Rekonstruktion. Die Wahl entscheidet wesentlich über die Qualität der berechneten Näherung und ist ein schwieriges – und weiterhin

aktuelles – Thema, das hier aber nicht weiter verfolgt werden soll.

Die fundamentale Frage ist hierbei, *welcher* Funktionenraum für die Beschreibung von Bildern geeignet ist. Hier stellen sich die üblichen Räume durchweg als unzureichend dar: Sollen die Funktionen lediglich integrierbar sein (d.h. Lebesgue-Funktionen), so enthalten sie auch solche, die von Rauschen nicht zu unterscheiden sind – sie taugen also nicht für das Entrauschen.

Stetige oder (schwach) differenzierbare Funktionen (d. h. Sobolev-Funktionen) lassen aber entweder keine Sprünge zu – taugen also nicht für die Beschreibung von Bildern mit scharfen Kanten, etwa zwischen Vorder- und Hintergrund – oder führen zu Minimierungsaufgaben, die beweisbar keine Lösung haben. Als "Goldlöckchen-Raum" hat sich stattdessen der Raum der sogenannten *Funktionen von beschränkter Variation* (BV) herausgestellt: Diese können Sprünge über Kanten enthalten, davon abgesehen besitzen sie aber genügend Regularitätseigenschaften, um rauschähnliches Verhalten auszuschließen. Der entsprechende Regularisierungsterm ist die *totale Variation* (TV), die bei stückweise konstanten Bildern den Umfang der Stücke mal der Höhe der Sprünge zwischen benachbarten Stücken misst. Die präzise Definition (über Dualität mit Testfunktionen) ist technisch, führt aber auf sehr interessante Fragestellungen der geometrischen Maßtheorie.[2] Für unsere Zwecke reicht aus, dass bei Betrachtung von *diskreten* (d.h. auf einem bestimmten Pixel-Gitter als stückweise konstant angenommenen) Bildern die totale Variation mit der 1-Norm des diskreten Gradienten (dem Vektor der Differenzen aller benachbarter Pixelwerte) übereinstimmt. Für Lösungen der entsprechenden Minimierungsaufgaben wird der diskrete Gradient also – analog zum compressed sensing – möglichst oft Null und damit die Lösung selber in der Regel stückweise konstant sein.

Eine ähnliche Frage stellt sich bei dem Rauschen. Aus der Antwort – welche Struktur von dem Rauschen erwartet werden kann – leitet sich der konkrete Diskrepanzterm ab; oft spielen hierbei statistische Überlegungen eine Rolle. So ist zum Beispiel bekannt, dass für Gaußsches (d.h. unabhängig normalverteiltes) Rauschen die Euklidische Distanz der Differenz von verrauschtem und gesuchtem Bild (als *maximum likelihood estimator*) am günstigsten ist. Da physikalische Prozesse oft zu einem entsprechenden thermischen Rauschen führen (und die Normalverteilung Universalitätseigenschaften hat), ist dieser Diskrepanzterm weit verbreitet. In der digitalen Bildgebung spielen aber noch weitere Prozesse eine Rolle, die nicht mehr normalverteilt sind.

Ein Beispiel ist *Impuls-Rauschen*, bei dem einzelne Pixel mit einer gewissen Wahrscheinlichkeit durch einen großen Zufallswert gestört werden, oder *Salz-und-Pfeffer-Rauschen*, bei dem einzelne Pixel entweder maximal gesättigt oder komplett schwarz werden. Ersteres kann durch defekte Signalleitungen oder Einfall kosmischer Strahlung entstehen, letzteres durch defekte Bildsensoren.

In beiden Fällen sind die nicht betroffenen Bildpixel exakt erhalten – das Rauschen ist also *sparse*, und ein geeigneterer Diskrepanz-Term ist die 1-Norm der Differenz von verrauschtem und gesuchtem Bild. Abbildung (1) zeigt beispielhaft die mit diesem Ansatz erzielbaren Ergebnisse; bei genauerem Hinsehen erkennt man die typische „cartoonartige" Lösung bei Minimierung der totalen Variation. Weitere Beispiele, die andere Diskrepanzterme benötigen, sind uniformes Rauschen (etwa durch Rundungsfehler bei der Quantisierung z.B. im Mobilfunk oder bei der JPEG-Kompression) oder Poisson-Rauschen (das etwa in der Astronomie oder der Fluoreszenz-Mikroskopie auftritt).

Variationsmethoden in der Bildverarbeitung führen also auf nichtglatte Optimierungsprobleme. Da Bilder immer höhere Auflösung haben, und jedes Pixel eine unbekannte Variable darstellt – ganz zu schweigen von der Anzahl von Unbekannten, wenn man Videos betrachtet – sind effiziente Verfahren für solche Probleme offensichtlich von großer Bedeutung; erst durch ihre Entwicklung haben die Variationsmethoden ihre praktische Relevanz entwickelt. Die Betrachtung von „idealisierten" Bildern als Funktionen anstelle von konkreten pixelbasierten Darstellungen führt dabei zu Verfahren, die optimal mit der Anzahl der Pixel skalieren.

Konvexe Analysis

Wie löst man nun solche nicht differenzierbaren Minimierungsprobleme? Ein naheliegender Ansatz wären sogenannte *ableitungsfreie Verfahren* aus der nichtlinearen Optimierung, die nur Funktionswerte in vorgegebenen oder geeignet ausgewählten Punkten verwenden. Beispiele sind das Nelder-Mead-Simplex-Verfahren (nicht zu verwechseln mit dem Simplex-Verfahren aus der linearen Optimierung) oder stochastische Suchverfahren wie *simulated annealing*. Diese basieren entweder auf einer lokalen Interpolation durch Polynome, die dann exakt minimiert werden können, oder auf mehr oder weniger strukturiertem Abtasten des gesamten Suchraums.

Entsprechend wenig kann man auch darüber aussagen, wann und wie gut diese Verfahren funktionieren; sie stellen eine Art letzter Ausweg dar für Probleme, in denen man gar keine Struktur findet. Für Probleme aus der Bildverarbeitung mit mehreren Millionen Unbekannten sind sie jedenfalls gänzlich ungeeignet.

Stattdessen sucht man nach *verallgemeinerten Ableitungsbegriffen*, die auch für nicht differenzierbare Funktionen im üblichen Sinn a) Minimierer als (verallgemeinerte) Nullstelle haben, und b) genug Rechenregeln zulassen, um für eine

(a) Originalbild.

(b) aufgenommenes Bild.

(1) Bildschärfen und Entfernen von Salz-und-Pfeffer-Rauschen mit 1-Norm und totaler Variation[4].
Quelle: Bilder aus Clason, Jin und Kunisch 2010

möglichst große Klasse von Funktionen explizit angegeben werden können. Die Hoffnung ist, dadurch das Vorgehen im differenzierbaren Fall weitestgehend übertragen zu können. Besonders gut funktioniert dies für *konvexe Funktionen*, zu denen die meisten in der Bildverarbeitung vorkommenden Funktionen gehören; entsprechend bezeichnet man ihr Studium als *konvexe Analysis*. Anschaulich ist eine skalare Funktion konvex, wenn jede ihrer Tangenten komplett unterhalb des Funktionsgraphen bleibt; die mathematische Definition im Fall einer reellwertigen Funktion auf einem beliebigen (normierten) Vektorraum bildet genau diese geometrische Anschauung im Wesentlichen ab.

Offensichtlich gilt dies für unser Beispiel der Betragsfunktion. Diese Anschauung treibt auch das weitere Vorgehen: geometrisch entspricht die Ableitung einer Funktion in einem Punkt der Steigung der Tangente in diesem Punkt; ist die Funktion nicht differenzierbar weil die Tangente – und damit deren Steigung – nicht eindeutig ist, definieren wir einfach das *Subdifferential* der Funktion in diesem Punkt als die *Menge aller* Tangentensteigungen. Existiert nun unter allen Tangenten in einem Punkt eine, die horizontal ist – ist also die Null in dem Subdifferential in diesem Punkt enthalten – so muss dieser Punkt ein Minimierer sein. Der Preis dafür, dass wir die differenzierbaren Funktionen verlassen haben, ist also, dass wir statt einer Ableitung stets mit einer Menge (die auch leer sein kann!) rechnen müssen. Die üblichen Rechenregeln wie die Summenregel (Ableitung der Summe ist Summe der Ableitungen) und Kettenregel gelten im Allgemeinen also nur als Inklusionen der beteiligten Mengen; um Gleichheit zu garantieren, müssen zusätzliche Voraussetzungen gelten.[5] Kann man diese nachweisen, erhält man aber für Probleme aus der Bildverarbeitung eine Charakterisierung mit Hilfe von Subdifferentialen des Diskrepanz- und Regularisierungsterms, die oft explizit dargestellt werden können. Für die Betragsfunktion ist dies zum Beispiel das *verallgemeinerte Vorzeichen*, das in der Null aus dem gesamten Intervall von -1 bis 1 besteht; das Subdifferential der 1-Norm ist komponentenweise gegeben durch das verallgemeinerte Vorzeichen der jeweiligen Komponenten.

Proximalpunkt- und Splitting-Verfahren

Diese verallgemeinerte Ableitung möchte man nun auch für praktisch umsetzbare Verfahren nutzen. Wieder ist der Ausgangspunkt das Vorgehen im differenzierbaren Fall; dort ist ein klassisches Verfahren das *Gradientenverfahren*, das in jedem Schritt in der aktuellen Näherung die negative Ableitung auswertet und dann eine geeignete Schrittweite lang in diese Richtung geht. Im konvexen Fall könnte man nun in der aktuellen Näherung ein Element aus dem Subdifferential (d.h. einen *Subgradienten*) auswählen und in dessen negative Richtung gehen. Dies funktioniert allerdings in der Regel nicht,

(c) rekonstruiertes Bild.

da das Subdifferential neben einer Abstiegsrichtung (d.h. einer, entlang der Funktionswert kleiner wird) auch beliebig viele Aufstiegsrichtungen enthalten kann. In den 1960er Jahren wurden daher sogenannte *Bundle-Verfahren* entwickelt, die aus mehreren Subgradienten auch zu zurückliegenden Schritten ein Modell der Funktion bauen und dann versuchen, dieses zu minimieren. Diese Verfahren funktionieren zwar, sind aber ebenfalls nicht effizient genug für die in der Bildverarbeitung relevanten Problemgrößen.

Es stellt sich nun heraus, dass anstelle eines expliziten Subgradientenschrittes ein impliziter Subgradientenschritt (analog zum Euler-Verfahren für gewöhnliche Differentialgleichungen) zu einem beweisbar konvergenten Verfahren führt. Die Idee besteht darin, die zu lösende Mengeninklusion „Null liegt im Subdifferential", die den gesuchten Minimierer charakterisiert, mit Hilfe einer sogenannten *Proximalpunktabbildung* (die der Auswertung des impliziten Subgradientenschrittes entspricht) als Fixpunkt*gleichung* umzuschreiben. Diese kann dann durch eine Fixpunktiteration gelöst werden; der Ansatz heißt daher auch *Proximalpunktverfahren*.

Allerdings ist die Auswertung der Proximalpunktabbildung in der Regel genauso kompliziert wie die Minimierung des ursprünglichen Funktionals, so dass das Verfahren nicht praktikabel ist. Wendet man diesen Ansatz jedoch auf die durch die konvexen Rechenregeln erhaltene Charakterisierung an, so sind nur die Proximalpunktabbildungen der separaten Funktionen auszuwerten – und diese haben oft eine explizite Darstellung. Für die Betragsfunktion erhält man zum Beispiel den sogenannten *soft-thresholding*-Operator, der alle Werte in einer Umgebung der Null auf Null setzt und die restlichen Werte um einen entsprechenden Betrag in Richtung Null „verschiebt", damit der Operator stetig bleibt. Die Proximalpunktabbildung der 1-Norm wendet einfach komponentenweises *soft-thresholding* an und kann daher massiv parallel (etwa auf Graphikkarten) für jedes Pixel eines Bildes separat berechnet werden. Durch ihre einfache Umsetzung haben sich diese *Splitting-Verfahren* rasch zu den Arbeitspferden der modernen Bildverarbeitung entwickelt; allen voran die unter dem Namen *Chambolle-Pock-Verfahren* bekannte Variante, die besonders gut für die Minimierung der totalen Variation geeignet ist.[6] Aktuelle Weiterentwicklungen enthalten zusätzlich Extrapolationsschritte mit geeignet gewählten Schrittlängen, die die Verfahren weiter beschleunigen. (Dies entspricht im Wesentlichen dem Schritt vom Gradientenverfahren zum Verfahren der konjugierten Gradienten (CG)).[7]

Semiglatte Newton-Verfahren

Auch mit Extrapolation sind oft sehr viele Schritte nötig, um dem gesuchten Minimierer für eine ansprechende Bildrekonstruktion hinreichend nahe zu kommen; dies ist insbesondere bei Problemen aus der medizinischen Bildgebung der Fall. Effizienter sind *Newton-Verfahren*, die deutlich weniger Schritte benötigen. Das klassische Newton-Verfahren für nichtlineare Gleichungen verwendet die Ableitung in jedem Schritt, um den nächsten Punkt als Nullstelle einer „optimalen" linearen Näherung – und damit explizit – berechnet. In der nichtlinearen Optimierung wird dies natürlich auf die Ableitung der zu minimierenden Funktion angewendet, deren Nullstelle der gesuchte Minimierer ist. Ist die Funktion selber nicht differenzierbar, scheint dieser Ansatz zunächst hoffnungslos. Für konvexe Funktionen kann man Minimierer aber wie im letzten Abschnitt beschrieben durch Fixpunkte der entsprechenden Proximalpunktabbildung charakterisieren – man erhält also doch wieder eine Gleichung, die allerdings selber nicht differenzierbar ist. Das Spiel geht also von vorne los: Gesucht wird ein (neuer) verallgemeinerter Ableitungsbegriff, der a) anstelle der klassischen Ableitung im Newton-Verfahren eingesetzt werden kann und b) für genügend viele praktisch relevante Funktionen ausgewertet werden kann. Für viele Proximalpunktabbildungen – darunter die der 1-Norm – stellt sich nun heraus, dass sie zumindest *stückweise* differenzierbar sind. Ist man nun in einem Newton-Schritt an einer Stelle, in der die Funktion differenzierbar ist, nimmt man einfach die klassische Ableitung; in den „Knicken", wo die Abbildung nicht differenzierbar ist, wird man schon nicht so oft landen, dass es einen Unterschied macht – dort nimmt man einfach die nächstbeste Ableitung. Dieser einfache Ansatz funktioniert tatsächlich; daraus eine rigorose Konvergenztheorie zu entwickeln, ist natürlich deutlich schwieriger und führt auf das Konzept der *Newton-Ableitung* und dem zugehörigen *semiglatten Newton-Verfahren*.[8] Trotzdem sind auch diese Verfahren leicht und effizient umzusetzen; sie wurden zum Beispiel verwendet, um die Ergebnisse in Abbildung (1) zu erzeugen.[9] Semiglatte Newton-Verfahren haben sich auch in vielen weiteren Problemen – nicht nur mit 1-Normen – aus Bildverarbeitung und optimaler Steuerung bewährt.[10]

Ausblick

Die aktuelle Forschung auf diesem Gebiet beschäftigt sich mit der Entwicklung von Verfahren, die auch für nichtkonvexe Funktionen funktionieren, etwa für die p-„Normen" mit p kleiner als 1, die an Stelle der 1-Norm in der totalen Variation oft zu einer besseren Rekonstruktion führen.

Auch in der *nichtglatten Analysis* sucht man nach weiter verallgemeinerten Ableitungskonzepten, die auch für nichtkonvexe Funktionen zur Charakterisierung von Minimierern eingesetzt werden können. Während dies für Lipschitz-stetige Funktionen mit dem Clarke-Subdifferential noch weitgehend gelungen ist, bleibt die allgemeine Theorie noch unbefriedigend – außer in Einzelfällen erfüllen die bislang untersuchten verallgemeinerten Ableitungen entweder nicht mehr alle gewünschten Rechenregeln exakt oder können nicht mehr explizit angegeben werden.

Es bleibt also noch genug zu tun.

Summary

Modern variational methods in mathematical imaging formulate the task of denoising, deblurring, or reconstructing images – photographic or medical – as an optimization problem in which the goal is to find the "optimal" image that minimizes a weighted sum of a discrepancy term, which measures the distance of a candidate image to the given data, and a regularization term, which measures the abstract "goodness" of the candidate image. For the latter, the total variation has turned out to be especially suitable for images, since it allows jumps in the minimizer (corresponding to sharp edges in the image) while still imposing sufficient regularity otherwise to remove noise. However, this term is not differentiable, and hence efficient nonsmooth optimization methods are needed. For convex functions such as the total variation, techniques of convex analysis make it possible to characterize minimizers as fixed points that can be computed either through fixed-point iteration – leading to proximal point and splitting methods – or through iterative linearization – leading to semismooth Newton methods. Current research focuses on generalizing these approaches to nonconvex functions.

Anmerkungen

1) Candès, Romberg und Tao 2006.
2) Attouch, Buttazzo und Michaille 2006; Valkonen 2014.
4) Bilder aus Clason, Jin und Kunisch 2010.
5) Erst hierfür wird tatsächlich die Konvexität der Funktion benötigt.
6) Chambolle und Pock 2011.
7) Beck 2017.
8) Ulbrich 2011; Ito und Kunisch 2008.
9) Clason, Jin und Kunisch 2010.
10) Zahlreiche Beispiele finden sich in den Publikationen des Autors, die auf https://www.uni-due.de/mathematik/agclason/clason_pub.php heruntergeladen werden können.

Literatur

– Attouch, H., Buttazzo, G. und Michaille, G. (2006), Variational Analysis in Sobolev and BV Spaces, Bd. 6, MPS/SIAM Series on Optimization, Society for Industrial und Applied Mathematics (SIAM), doi: 10.1137/1.9780898718782.
– Beck, A. (2017), First-Order Methods in Optimization, Philadelphia, PA: Society for Industrial und Applied Mathematics, doi: 10.1137/1.9781611974997.
– Bredies, K. und Lorenz, D. A. (2011), Mathematische Bildverarbeitung, Vieweg+Teubner, doi: 10.1007/978-3-8348-9814-2.
– Candès, E. J., Romberg, J. K. und Tao, T. (2006), Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59.8, S. 1207–1223, doi: 10.1002/cpa.20124.
– Chambolle, A. und Pock, T. (2011), A first-order primal-dual algorithm for convex problems with applications to imaging, J Math Imaging Vis 40.1, S. 120–145, doi: 10.1007/s10851-010-0251-1.
– Clason, C. (2017), Nonsmooth Analysis and Optimization, Vorlesungsskript.
– Clason, C., Jin, B. und Kunisch, K. (2010), A duality-based splitting method for 1–TV image restoration with automatic regularization parameter choice, SIAM Journal on Scientific Computing 32.3, S. 1484–1505, doi: 10.1137/090768217.
– Hiriart-Urruty, J.-B. und Lemaréchal, C. (2001), Fundamentals of Convex Analysis, Berlin: Springer-Verlag, doi: 10.1007/978-3-642-56468-0.
– Ito, K. und Kunisch, K. (2008), Lagrange Multiplier Approach to Variational Problems and Applications, Bd. 15, Advances in Design and Control, SIAM, doi: 10.1137/1.9780898718614.
– Ulbrich, M. (2011), Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces, Bd. 11, MOS-SIAM Series on Optimization, Philadelphia, PA: SIAM, doi: 10.1137/1.9781611970692.
– Valkonen, T. (2014), Measure and Image, Vorlesungsskript, University of Cambridge, url: http://tuomov.iki.fi/mathematics/mandi.pdf.

Der Autor

Christian Clason, geboren 1974, studierte Mathematik an der Ludwig-Maximilians-Universität München. Nach der Promotion an der Technischen Universität München (2006) wechselte er an die Karl-Franzens-Universität Graz, wo er sich 2012 habilitierte. Seit 2014 ist er Professor für inverse Probleme an der Fakultät für Mathematik der Universität Duisburg-Essen. Seine Forschung beschäftigt sich mit nichtglatter Analysis und Optimierung in inversen Problemen, der optimalen Steuerung von partiellen Differentialgleichungen und der biomedizinischen Bildgebung. Sie liegt damit an der Schnittstelle von Funktionalanalysis, Optimierung und Numerik.