Titel/Title: A Generalized Conditional Gradient Method and its Connection to an Iterative Shrinkage Method

Autor*innen/Author(s): Kristian Bredies, Dirk A. Lorenz, Peter Maass

Veröffentlichungsversion/Published version: Postprint

Publikationsform/Type of publication: Artikel/Aufsatz

Empfohlene Zitierung/Recommended citation:

Bredies, K., Lorenz, D.A. & Maass, P. A generalized conditional gradient method and its connection to an iterative shrinkage method. Comput Optim Appl 42, 173–193 (2009). https://doi.org/10.1007/s10589-007-9083-3

Verfügbar unter/Available at:
(wenn vorhanden, bitte den DOI angeben/please provide the DOI if available)

https://doi.org/10.1007/s10589-007-9083-3

Zusätzliche Informationen/Additional information:

This article combines techniques from two fields of applied mathematics: optimization theory and inverse problems. We investigate a generalized conditional gradient method and its connection to an iterative shrinkage method, which has been recently proposed for solving inverse problems.

A Generalized Conditional Gradient Method and its Connection to an Iterative Shrinkage Method

Kristian Bredies, Dirk A. Lorenz, Peter Maass
Fachbereich 03
Universität Bremen
Postfach 33 04 40
28334 Bremen
Germany

{kbredies,dlorenz,pmaass}@math.uni-bremen.de

September 23, 2008

Abstract

This article combines techniques from two fields of applied mathematics: optimization theory and inverse problems. We investigate a generalized conditional gradient method and its connection to an iterative shrinkage method, which has been recently proposed for solving inverse problems.

The iterative shrinkage method aims at the solution of non-quadratic minimization problems where the solution is expected to have a sparse representation in a known basis. We show that it can be interpreted as a generalized conditional gradient method. We prove the convergence of this generalized method for general class of functionals, which includes non-convex functionals. This also gives a deeper understanding of the iterative shrinkage method.

1 Introduction

Optimization theory and the theory of inverse problems are both well-developed fields of applied mathematics, which analyze similar problems and related algorithms. For example a large class of methods from inverse problems can be interpreted as minimization algorithms for convex functionals and thus are, in turn well-known techniques for solving optimization problems. The present paper combines techniques from both fields as it investigates classical optimization algorithms and their counterparts in the theory of inverse problems.

We present two main results: The first result shows, that the recently proposed iterative shrinkage method for solving inverse problems [3] can be interpreted as a special case of a generalization of the well-known conditional gradient method for constrained minimization problems. As a second result we analyze the convergence properties of our proposed generalized conditional gradient method. The convergence properties also hold for a certain class of non-convex functionals.

The article is organized as follows: In Section 2 we briefly review conditional gradient methods and propose a natural generalization. The convergence properties of this generalized gradient method are analyzed in Section 3. The range of applicability of this convergence result will include the minimization of non-convex functionals.

We then review the iterative shrinkage method as it is used in the theory of inverse problems. Section 4 is devoted to analyze the connection between the iterative shrinkage approach via surrogate functionals and generalized conditional gradient methods. As a result we also obtain a new proof for the convergence of the iterative shrinkage method.

2 Iterative Gradient Methods

Typical problems in optimal control theory ask to minimize a (generally non-convex) functional F with restrictions given by a convex, closed and non-empty set U_{ad} in a Hilbert space H:

$$\min_{u \in U_{\mathrm{ad}}} F(u). \tag{1}$$

This type of problem is well-understood regarding existence and uniqueness of a solution [16] as well as the convergence of several approximate computational methods. A large number of such algorithms has been analyzed in the 1970's already, see [13] and the references therein. One of these methods, originally proposed by Frank and Wolfe [9], is the well known conditional gradient method. As it is typical for gradient methods it is easy to apply in most cases but it exhibits poor convergence rates.

2.1 The Conditional Gradient Method

For the sake of completeness, we start with a short description of the conditional gradient method. We also review some well-known facts of this method.

Assume, that F is Gâteaux differentiable and the set of admissible vectors U_{ad} is bounded. The conditional gradient method generates a sequence $\{u_n\}$ of approximations according to the following procedure.

1. Choose $u_0 \in U_{\mathrm{ad}}$ arbitrarily, $n = 0$
2. Determine an approximate solution v_n by solving
$$\min_{v \in U_{\mathrm{ad}}} \langle F'(u_n), v \rangle \ .$$
3. Perform a line-search with direction $v_n - u_n$ by solving
$$\min_{s \in [0,1]} F\big(u_n + s(v_n - u_n)\big) \ ,$$
denote s_n as a solution of this problem.

4. Set $u_{n+1} = u_n + s_n(v_n - u_n)$ and $n = n + 1$, return to Step 2.

Note, since v_n is a solution of the linearized minimization problem, first order necessary conditions are fulfilled as soon as $\langle F'(u_n), v_n - u_n \rangle \geq 0$. Thus the method is suitable for finding stationary points of F. Moreover, because U_{ad} is bounded, each iterate is well-defined.

In general, the solution of the minimization problems in Step 2 and Step 3 is not unique, the algorithm only requires to choose one of the minimizers. In each step, the procedure involves solving a linear constrained minimization problem as well as a one-dimensional optimization problem. Usually, this is considered to be easily done, as illustrated by the following constrained least squares example.

Example 1. Let $\Omega \subset \mathbf{R}^d$ be a bounded domain and define $H = \mathrm{L}^2(\Omega)$. Let $U_{\mathrm{ad}} = \{u : |u| \leq 1\}$ denote the set of admissible functions and define

$$F(u) = \frac{\|Ku - f\|^2}{2} \quad ,$$

with a continuous linear operator $K : H \to H$ and $f \in H$. In this case, the conditional gradient method generates a sequence of approximations $\{u_n\}$ by

$$p_n = K^*(Ku_n - f) \quad , \qquad v_n(x) = \begin{cases} -1 & p_n(x) > 0 \\ 1 & p_n(x) < 0 \\ 0 & p_n(x) = 0 \end{cases}$$

$$s_n = P_{[0,1]}\left(\frac{\langle Ku_n - f, K(u_n - v_n) \rangle}{\|K(v_n - u_n)\|^2} \right) \quad , \qquad u_{n+1} = u_n + s_n(v_n - u_n) \quad ,$$

where $P_{[0,1]}$ denotes the projection onto the interval $[0, 1]$.

The convergence of this sequence $\{u_n\}$ to a solution of the problem (1) essentially requires that F is convex and the Fréchet derivative F' is Lipschitz continuous. Moreover, convergence rates can be established and – based on additional information on the solution – the convergence speed can be estimated, see e. g. [4, 5]. For non-convex optimization problems the situation is somewhat different. However, for compact U_{ad} the conditional gradient method still yields subsequences converging to a stationary point of F, see [6].

2.2 A Generalized Conditional Gradient Method

We will now motivate a generalization of the conditional gradient algorithm, which shares many convenient properties with the classical conditional gradient method. As we will see in Section 4, the proposed method also generalizes methods for non-quadratic regularization functionals in the theory of inverse problems.

For motivating the generalized method we return to problem (1). One notices that this constrained problem can actually be written as an "unconstrained" one with the help of the indicator functional

$$I_{U_{\text{ad}}}(u) = \begin{cases} 0 & u \in U_{\text{ad}} \\ \infty & u \notin U_{\text{ad}} \end{cases} .$$

Problem (1) thus can be reformulated as

$$\min_{u \in H} \; F(u) + \Phi(u)$$

where $\Phi = I_{U_{\text{ad}}}$.
To illustrate the proposed generalization, we summarize the key properties of F and Φ:

1. F is smooth (Gâteaux differentiable), but Φ is allowed to be non-differentiable.

2. The minimization of Φ is considered be solved easily. The minimization of F may be hard, in particular we will not assume convexity for F in the following.

With these assumptions in mind, we propose to analyze gradient methods for solving

$$\min_{u \in H} \; F(u) + \Phi(u) \quad , \tag{2}$$

where F is assumed to be Gâteaux differentiable and Φ is assumed to be proper, convex, lower semi-continuous and coercive with respect to the norm in H:

Condition 1. *Let the functional* $\Phi : H \to \,]-\infty, \infty]$ *satisfy:*

1. $\Phi(u) < \infty$ *for a* $u \in H$*,*

2. $\Phi\big(su + (1-s)v\big) \leq s\Phi(u) + (1-s)\Phi(v)$ *for all* $u, v \in H$ *and* $s \in [0,1]$*,*

3. $\Phi(u) \leq \liminf_{n\to\infty} \Phi(u_n)$ *whenever* $u = \lim_{n\to\infty} u_n$ *in* H*,*

4. $\Phi(u)/\|u\| \to \infty$ *whenever* $\|u\| \to \infty$*.*

The interesting cases covered by this approach are functionals where F is not convex and Φ is not differentiable.

Each gradient method first of all requires to determine descent directions. Unfortunately, the non-differentiability of Φ does not permit to use the gradient of the full functional in (2). Alternatively, we propose to choose a direction by

$$\min_{v \in H} \; \langle F'(u), \, v \rangle + \Phi(v) \; .$$

We note, that this reduces to the standard choice of direction for conditional gradient methods if $\Phi = I_{U_{\text{ad}}}$.

Hence, the "generalized conditional gradient method" proceeds as follows.

1. Choose $u_0 \in H$, such that $\Phi(u_0) < \infty$ and set $n = 0$.
2. Determine a solution v_n of

$$\min_{v \in H} \langle F'(u_n), v \rangle + \Phi(v) \; . \tag{3}$$

3. Set s_n as a solution of

$$\min_{s \in [0,1]} F\big(u_n + s(v_n - u_n)\big) + \Phi\big(u_n + s(v_n - u_n)\big) \; .$$

4. Put $u_{n+1} = u_n + s_n(v_n - u_n)$ and $n = n + 1$, return to Step 2.

The differentiability of F and the assumptions on Φ ensure that the functional in (3) is proper, convex and lower semi-continuous. Hence, standard arguments from convex analysis yield the existence of a minimizer in Step 2 of the algorithm, see [7].

We are going to analyze this generalized conditional gradient method in Section 3. There we show that, under the same conditions as above, the minimizing functional decreases in every step and moreover we establish criteria under which convergence can be ensured.

However, before we present the convergence analysis, we now review the iterative shrinkage method by means of surrogate functionals. Our main result in Section 4 will be an interpretation of this method as a generalized gradient method.

2.3 The Iterative Shrinkage Method

In the following, we briefly motivate and derive the iterative shrinkage algorithm as presented in [3]. The reader familiar with the method or more interested in the convergence result for the generalized conditional gradient method may skip this section.

In the theory of linear inverse problems, minimization problems are considered which often have a similar structure to the one presented in Example 1. An inverse problem deals with solving operator equations of the form

$$Ku = f$$

where f plays the role of measured data and K often models a kind of observation operator. Since these observation operators are often compact the solution of the inverse problem becomes ill-posed in the sense that the pseudoinverse of K is unbounded. A standard way for regularizing ill-posed problems is to minimize a modified discrepancy functional like

$$\|Ku - f\|^2 + \Phi(u).$$

For example, $\Phi(u) = \lambda \|u\|^2$ yields the well-known Tikhonov regularization.

In Example 1 the constraint $u \in U_{\mathrm{ad}}$ can be interpreted as a regularization of the ill-posed problem. Other penalization functionals are also widely used and often adapted to an a priori known smoothness of the solution [8, 12]. In [3], a class of regularization functionals is used which shall promote a sparse representation of the solution in a given basis. The authors consider the following minimization problem:

$$\min_{u\in H} \frac{\|Ku-f\|^2}{2} + \sum_n w_n |\langle u,\, \varphi_n\rangle|^p \tag{4}$$

where H is a separable Hilbert space, $K : H \to H$ a linear and continuous operator with $\|K\| < 1$, $\{\varphi_n\}$ an orthonormal basis of H, $\{w_n\}$ a non-negative sequence of weights and $p \geq 1$ an exponent.

The solution of this non-quadratic problem is straightforward, if the operator K is diagonal with respect to the basis $\{\varphi_n\}$. In this case the minimization problem reduces to one-dimensional minimization problems for each expansion coefficient. However, the minimization is not straightforward if K is not diagonal with respect to the chosen basis. In [3] an iterative approach based on a surrogate functional is proposed. There the authors assume $\|K\| < 1$, which can always be achieved after rescaling, and they define a surrogate functional by

$$\Psi_{w,p}(u,a) = \frac{\|Ku-f\|^2}{2} + \sum_n w_n|\langle u,\, \varphi_n\rangle|^p - \frac{\|Ku-Ka\|^2}{2} + \frac{\|u-a\|^2}{2} \ .$$

This surrogate functional has several nice properties:

- For $a = u$ the surrogate functional reduces to the original functional,
- the quadratic term $\|Ku\|^2$ cancels in the surrogate functional,
- for a fixed u minimizing $\Psi_{w,p}(u,a)$ with respect to a yields $a = u$,
- Minimizing $\Psi_{w,p}(u,a)$ with respect to u for a fixed a is equivalent to

$$\min_{u\in H} \sum_n w_n|\langle u,\, \varphi_n\rangle|^p + \tfrac{1}{2}|\langle a - K^*(Ka-f) - u,\, \varphi_n\rangle|^2 \ . \tag{5}$$

The minimizer is obtained by a pointwise shrinkage procedure. This can be done analytically by

$$u = \mathbf{S}_{w,p}\big(a - K^*(Ka-f)\big) = \sum_n S_{w_n,p}\big(\langle a - K^*(Ka-f),\, \varphi_n\rangle\big)\varphi_n \ , \tag{6}$$

where $S_{w,p}$ is a “shrinkage function” given by

$$S_{w,p}(x) = \begin{cases} \operatorname{sgn}(x)\big[|x| - w\big]_+ & p = 1 \\ G_{w,p}^{-1}(x) & p > 1 \end{cases}$$

with $G_{w,p}(x) = x + wp\,\operatorname{sgn}(x)|x|^{p-1}$.

Those observations lead to an efficient, iterative algorithm for the solution of (4): Take an arbitrary $u_0 \in H$, set $a_0 = u_0$ and minimize the corresponding surrogate functionals alternatingly with respect to the first and second argument. Since the solution of both problems are known, this leads to successive application of $\mathbf{S}_{w,p}$:

$$u_{n+1} = \mathbf{S}_{w,p}(u_n - K^*(Ku_n - f)).$$

The main result in [3] proves that the sequence $\{u_n\}$ generated by this algorithm converges to the minimizer of (4) under the conditions that

1. the exponent satisfies $1 \leq p \leq 2$,

2. the weights w_n are uniformly bounded away from zero.

3 Convergence Analysis

In this section we analyze the convergence properties of the generalized conditional gradient method for minimizing functionals of type (2)

$$\min_{u \in H} \; F(u) + \Phi(u) \quad .$$

The convergence results will include the case of non-convex functionals F.

3.1 Basic Descent Properties

One of the key ingredients for the analysis of any descent method is the first order necessary condition for a minimizer of (2). The following statement is commonly known and can, for example, be found in [11].

Lemma 1. *Let $F : H \to \mathbf{R}$ denote a Gâteaux-differentiable functional and assume that Φ satisfies Condition 1. Then, the first order necessary condition for optimality with respect to* (2) *is given by*

$$u \in H \; : \quad \langle F'(u), \, v - u \rangle \geq \Phi(u) - \Phi(v) \quad \text{for all } v \in H \; .$$

This condition is equivalent to

$$\langle F'(u), \, u \rangle + \Phi(u) = \min_{v \in H} \; \langle F'(u), \, v \rangle + \Phi(v) \; . \tag{7}$$

Proof. Let u be a solution of the minimization problem (2). For an arbitrary $v \in H$ and $s \in \,]0, 1]$ we have by the convexity of Φ

$$\begin{aligned} F(u) + \Phi(u) &\leq F\big(u + s(v - u)\big) + \Phi\big(u + s(v - u)\big) \\ &\leq F\big(u + s(v - u)\big) + (1 - s)\Phi(u) + s\Phi(v) \; . \end{aligned}$$

Reordering the inequality and dividing by s gives

$$\Phi(u) - \Phi(v) \le \frac{F\big(u + s(v-u)\big) - F(u)}{s} \quad .$$

Taking the limit $s \to 0$ yields the desired necessary condition. The second characterization follows immediately. $\square$

The identity (7) reveals that the first order necessary condition is connected to the descent direction of the generalized conditional gradient algorithm: the descent direction is obtained by minimizing the right hand side of (7) in each step of the algorithm. As a direct consequence of the first order condition we see that a stationary point is reached as soon as $\langle F'(u_n), v_n - u_n\rangle \ge \Phi(u_n) - \Phi(v_n)$, which provides a test for stationarity.

On the other hand, if a stationary point is not reached, a reasonable method should determine a descent direction such that the functional (strictly) decreases in the next step. This is ensured by the following lemma.

Lemma 2. *Let F and Φ be defined as in Lemma 1. Suppose u_n does not fulfill the first order necessary conditions. Then the generalized conditional gradient method determines a u_{n+1} such that*

$$F(u_{n+1}) + \Phi(u_{n+1}) < F(u_n) + \Phi(u_n) \ .$$

Proof. Denote v_n as the solution of the linearized minimization problem according to Step 2 of the algorithm. Since the first order necessary conditions are not fulfilled, there is a $c_n > 0$ such that

$$-\langle F'(u_n), v_n - u_n\rangle - c_n = \Phi(v_n) - \Phi(u_n) \ .$$

The convexity of Φ then yields

$$\begin{aligned}\Phi\big(u_n + s(v_n - u_n)\big) &\le \Phi(u_n) + s\big(\Phi(v_n) - \Phi(u_n)\big) \\ &= \Phi(u_n) - s\langle F'(u_n), v_n - u_n\rangle - sc_n \ .\end{aligned}$$

Exploiting the Gâteaux-differentiability of F, we choose an $0 < \varepsilon < c_n$ such that

$$F\big(u_n + s(v_n - u_n)\big) = F(u_n) + s\langle F'(u_n), v_n - u_n\rangle + r(s)$$

with $|r(s)| \le \varepsilon s$ for $0 \le s \le \delta$ where $1 \ge \delta > 0$. Combining this with the above inequality yields

$$(F + \Phi)\big(u_n + s(v_n - u_n)\big) \le (F+\Phi)(u_n) - (c_n - \varepsilon)s \ .$$

This ensures the decrease of the functional $F + \Phi$ for sufficiently small s. Since s_n is obtained by minimizing over all $s \in [0,1]$, this also has to be true for s_n. $\square$

Remark 1. The proof also shows that it is not necessary to choose the step size s optimally. The functional decreases as soon as $0 < s < \delta$. So, if one is able to control the derivative of F, one can actually calculate δ and obtain a feasible interval for s. This might be useful if the line-search problem cannot be solved explicitly.

For the classical conditional gradient algorithm, this would lead to an explicit step size rule which requires a Lipschitz continuous Fréchet derivative of the minimizing functional on the feasible set U_{ad}. Such rules usually contain the solution of a line minimization problem with a quadratic term [4]. A generalization of this to the generalized conditional gradient algorithm would lead to a line minimization problem of a quadratic term plus the functional Φ (or at least a reasonable estimation). This is practically the same difficulty as performing a direct line-search, thus the gain is relatively small.

Remark 2. A slightly different approach is based on the use of implicitly defined step sizes, which e. g. were proposed by Armijo [1] and Goldstein [10]. They provide a sufficiently large descent of the functional and can easily be adapted to the generalized conditional gradient algorithm. For instance, the Armijo scheme would choose $s = \beta^k$ with $\beta \in]0, 1[$ and k the smallest non-negative integer satisfying

$$\alpha\beta^k \big(\langle F'(u),\, u - v\rangle + \Phi(u) - \Phi(v)\big) \leq (F + \Phi)(u) - (F + \Phi)(u + \beta^k(v - u))$$

with $\alpha \in]0, \frac{1}{2}[$. This is always possible if F is continuously Fréchet differentiable. To see this, we introduce the function

$$W(s) = \frac{(F + \Phi)(u) - (F + \Phi)(u + s(v - u))}{s\big(\langle F'(u),\, u - v\rangle + \Phi(u) - \Phi(v)\big)}$$

and investigate its behavior as s approaches zero. Since F is continuously differentiable and Φ is convex, one can use the intermediate value theorem and estimate

$$\begin{aligned} W(s) &= \frac{s\langle F'(\xi),\, u - v\rangle + \Phi(u) - \Phi(u + s(v - u))}{s\big(\langle F'(u),\, u - v\rangle + \Phi(u) - \Phi(v)\big)} \\ &\geq \frac{s\big(\langle F'(\xi),\, u - v\rangle + \Phi(u) - \Phi(v)\big)}{s\big(\langle F'(u),\, u - v\rangle + \Phi(u) - \Phi(v)\big)} = \frac{\langle F'(\xi),\, u - v\rangle + \Phi(u) - \Phi(v)}{\langle F'(u),\, u - v\rangle + \Phi(u) - \Phi(v)} \end{aligned}$$

where ξ is a point on the line connecting u and $u + s(v - u)$. Hence, the right hand side tends to 1 if $s \to 0$. Thus for small s, $W(s)$ becomes greater than any $\alpha \in]0, \frac{1}{2}[$. So, one can always choose a step size as desired.

This type of step size rules always yield algorithms having the same structure as the generalized conditional gradient algorithm and can be discussed analogously. However, our focus is on showing the connection between methods used for inverse problems and optimal control. In order to present how these results are connected, we always choose the generalized conditional gradient method with an optimal line-search.

3.2 Convergence for Non-Convex F

In this section, we will investigate the minimization problem (2) where F is a non-convex but continuously Fréchet differentiable functional. In particular, we will analyze the convergence properties of the generalized conditional gradient method associated with such problems. Such situations may occur even if the original problem itself is convex, for example see Section 4 and the analysis of the surrogate functional approach (9).

We analyze the existence of a minimizer as well as the convergence to a stationary point. This is well known for the conditional gradient method if the set of admissible vectors U_{ad} is compact, see [6]. An analogous condition for the generalized version would be the compactness of the sets $\{u : \Phi(u) \leq t\}$ for every $t \in \mathbf{R}$. The existence of a minimizer is then immediate.

Lemma 3. *We assume the sum of functionals in* (2) *to be coercive, i. e.* $(F+\Phi)(u) \to \infty$ *when* $\|u\| \to \infty$. *Let* Φ *satisfy Condition 1 and assume*

$$E_t = \{u \in H \ : \ \Phi(u) \leq t\} \text{ to be compact for every } t \in \mathbf{R} \ .$$

If F *is bounded on bounded sets and lower semi-continuous, then there exists a minimizer* $u \in H$ *of* (2) *and the minimum of* F *is finite.*

Proof. First choose a minimizing sequence $\{u_n\}$ in H, i. e. $\lim_{n\to\infty} F(u_n) + \Phi(u_n) = \inf_{u\in H} F(u) + \Phi(u)$, where the limit $-\infty$ is allowed. $(F + \Phi)(u_n) \leq C_1$ and due to the coercivity, $\|u_n\| \leq C_2$. The boundedness condition of F gives $|F(u_n)| \leq C_3$ which implies $\Phi(u_n) \leq C_1 - C_3$. Hence, $u_n \in E_{C_1-C_3}$ for all $n \geq 0$.

Thus, a subsequence (also denoted by $\{u_n\}$) which converges to some $u \in H$ can be chosen. Note that Condition 1 on Φ implies that this functional is bounded from below by some $C_4 \in \mathbf{R}$, so $(F + \Phi)(u_n) \geq C_4 - C_3$, i. e. the infimum is actually finite. Finally,

$$F(u) + \Phi(u) \leq \liminf_{n\to\infty}(F + \Phi)(u_n) = \inf_{u\in H}(F + \Phi)(u) \ .$$

□

In the non-convex setting it is harder to impose conditions which ensure that the set of stationary points consists of only one point. We can only expect that the cluster points of the sequence generated by the generalized conditional gradient method are stationary points. Moreover, these stationary points do not necessarily achieve the global minimum of $F + \Phi$. Nevertheless, under mild assumptions one can get statements regarding convergence.

For the analysis of the limit points generated by the generalized conditional gradient method, we introduce the functional

$$\Psi(u) = \langle F'(u), u\rangle + \Phi(u) - \min_{v\in H} \big(\langle F'(u), v\rangle + \Phi(v)\big) \ , \tag{8}$$

which is non-negative, see Lemma 1, and has its zeros exactly at the stationary points of $F + \Phi$. As can be seen in Lemma 2, this functional also indicates the decay of the functional $F + \Phi$ in each step of the iteration (it coincides with the c_n introduced there).

Lemma 4. *Let* Φ *satisfy Condition 1 and assume* F *to be continuously Fréchet differentiable. Then* Ψ *in* (8) *is lower semi-continuous.*

Proof. We follow the line of proof in [6]. Let $\{u_n\}$ be given such that $\lim_{n\to\infty} u_n = u$ in H. For a $w \in H$ with $\Phi(w) < \infty$ we obtain

$$\begin{aligned}\Psi(u_n) &= \langle F'(u_n),\, u_n\rangle + \Phi(u_n) - \big(\min_{v\in H} \langle F'(u_n),\, v\rangle + \Phi(v)\big) \\ &\geq \langle F'(u_n),\, u_n - w\rangle + \Phi(u_n) - \Phi(w) \\ &= \langle F'(u_n) - F'(u),\, u_n - w\rangle + \langle F'(u),\, u_n - w\rangle + \Phi(u_n) - \Phi(w)\ .\end{aligned}$$

Since F' is continuous, the inner products converge, hence

$$\liminf_{n\to\infty} \Psi(u_n) \geq \langle F'(u),\, u - w\rangle + \Phi(u) - \Phi(w)$$

and taking the supremum over all w with $\Phi(w) < \infty$ yields

$$\liminf_{n\to\infty} \Psi(u_n) \geq \langle F'(u),\, u\rangle + \Phi(u) - \min_{v\in H} \big(\langle F'(u),\, v\rangle + \Phi(v)\big) = \Psi(u)\ .$$

□

As an immediate consequence of this lower semi-continuity we conclude, that the set of stationary points is closed.

We now turn to analyzing the convergence of the generalized conditional gradient method. As we have already remarked, the stationary points of $F+\Phi$ are characterized by $\Psi(u) = 0$. Hence, we want to show, that the sequence $\{u_n\}$ generated by the generalized conditional gradient method satisfies $\lim_{n\to\infty} \Psi(u_n) = 0$. This will imply, that all cluster points of this sequence are stationary points.

Lemma 5. *Let F and Φ satisfy the conditions of Lemma 4. Furthermore, assume that the sum $F + \Phi$ is coercive and F' is uniformly continuous on bounded sets. Then every sequence $\{u_n\}$ generated by the generalized conditional gradient method satisfies* $\lim_{n\to\infty} \Psi(u_n) = 0$.

Proof. First note that the uniform continuity of F' on bounded sets implies the boundedness of F' on bounded sets: Let $R > 0$ and choose a $0 < \delta < R$ such that whenever $v, w \in B_R(0)$ and $\|v - w\| \leq \delta$ there follows $\|F'(v) - F'(w)\| \leq 1$. Now choose an $N \in \mathbf{N}$ with $N \geq R/\delta$ and an arbitrary $w \in B_R(0)$. Denote by $w_n = \frac{n}{N} w$ and estimate

$$\|F'(w) - F'(0)\| \leq \sum_{n=1}^{N} \|F'(w_n) - F'(w_{n-1})\| \leq N$$

since $\|w_n - w_{n-1}\| \leq \delta$. This implies the boundedness of F'. Additionally, one also gets the boundedness of F.

Now since $(F + \Phi)(u_{n+1}) \leq (F + \Phi)(u_n)$ by Lemma 2, $F + \Phi$ is bounded from below, see Lemma 3, and $\Phi(u_0) < \infty$, there exists an $z \geq \min_{u\in H} F(u) + \Phi(u)$ such that $\lim_{n\to\infty} F(u_n) + \Phi(u_n) = z$. Moreover, due to the coercivity of $F + \Phi$, $\|u_n\| \leq C_1$ for a $C_1 > 0$ and all $n \geq 0$. Additionally, the mapping F' is bounded on bounded sets, hence $\{F'(u_n)\}$ is also bounded.

Now consider the (set-valued) operator taking a $u \in H$ to $v \in H$ such that

$$\langle u, v \rangle + \Phi(v) = \min_{w \in H} \big(\langle u, w \rangle + \Phi(w) \big) \ .$$

Since Φ fulfills Condition 1, such operators are bounded with respect to the input u, see e. g. [2]. Consequently, there exists a $C_2 > 0$ such that $\|v_n\| \leq C_2$ for any choice of $\{v_n\}$.

We now turn to estimating the descent of the functional itself. The minimizing property of the line-search for s and the intermediate value theorem yield

$$\begin{aligned}
&(F + \Phi)(u_{n+1}) - (F + \Phi)(u_n) \\
&\qquad \leq F\big(u_n + s(v_n - u_n)\big) - F(u_n) + \Phi\big(u_n + s(v_n - u_n)\big) - \Phi(u_n) \\
&\qquad \leq F\big(u_n + s(v_n - u_n)\big) - F(u_n) + s\big(\Phi(v_n) - \Phi(u_n)\big) \\
&\qquad \leq -s\Psi(u_n) + s\langle F'(w_n(s)) - F'(u_n), v_n - u_n \rangle
\end{aligned}$$

where $w_n(s) = u_n + ts(v_n - u_n)$ for some $t \in [0, 1]$. Further, we obtain

$$\Psi(u_n) \leq \frac{(F + \Phi)(u_n) - (F + \Phi)(u_{n+1})}{s} + \|F'(w_n(s)) - F'(u_n)\|(C_1 + C_2) \ .$$

F' is uniformly continuous on bounded sets, this implies that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|w - u_n\| < \delta$ implies $\|F'(w) - F'(u_n)\| < \varepsilon/(2(C_1 + C_2))$. For $0 < s < \delta/(C_1 + C_2)$, this yields $\|w_n(s) - u_n\| \leq s\|v_n - u_n\| < \delta$ for every $n \geq 0$ and

$$0 \leq \Psi(u_n) < \frac{(F + \Phi)(u_n) - (F + \Phi)(u_{n+1})}{s} + \frac{\varepsilon}{2} \ .$$

Since $(F+\Phi)(u_n)$ converges, there is an index n_0 for which $(F+\Phi)(u_n) - (F+\Phi)(u_{n+1}) < \varepsilon s/2$ if $n \geq n_0$, thus $0 \leq \Psi(u_n) < \varepsilon$. □

We can now continue in two directions: we either aim at results on weak convergence, this would require at least a weakly lower semi-continuous Ψ in Lemma 4, or we strengthen the assumption on the sets E_t. We will follow the second approach and assume

$$E_t = \{u \in H \ : \ \Phi(u) \leq t\} \ \text{ is compact for every } \ t \in \mathbf{R} \ .$$

Note that the compactness condition also implies the existence of a convergent subsequence of $\{u_n\}$ whose limit has to be a stationary point. Together, both lemmata give the following convergence theorem.

Theorem 1. *Let Φ satisfy Condition 1 and assume every $E_t = \{u \in H \ : \ \Phi(u) \leq t\}$ to be compact for every $t \in \mathbf{R}$. Furthermore, let F be a continuously Fréchet differentiable functional which is bounded on bounded sets with $F + \Phi$ coercive, assume that $u_0 \in H$ obeys $\Phi(u_0) < \infty$. Let $\{u_n\}$ denote the sequence generated by the generalized conditional gradient method.*

Then there exists a convergent subsequence of $\{u_n\}$ and every convergent subsequence of $\{u_n\}$ converges to a stationary point of the functional $F + \Phi$.

Proof. We only have to show that the proof of Lemma 5 can be modified such that the statement remains true if uniform continuity of F' on bounded sets can be replaced with the compactness assumption on E_t together with the boundedness of F. For this, we will prove that all u_n and v_n are contained in a convex compact set on which F' is uniformly continuous.

Since F is bounded on bounded sets, the functional values $(F+\Phi)(u_n)$ still converge to a finite $z \in \mathbf{R}$ and also $\Phi(u_n) \leq C_1$ which means that the sequence $\{u_n\}$ is contained in a convex compact set. Furthermore, $\{F'(u_n)\}$ is bounded since it is contained in the continuous image of a compact set. As can be seen in the proof of Lemma 5, $\{v_n\}$ is bounded as well. Due to the minimizing property of v_n there is

$$\Phi(v_n) \leq \langle F'(u_n),\, u_n - v_n \rangle + \Phi(u_n) \leq C_2 + C_1$$

with a suitable $C_2 > 0$, which means that u_n, v_n as well as the lines connecting them are contained in the compact set $E_{C_1+C_2}$. □

We want to emphasize, that this theorem covers the case of non-convex functionals F. We can state a stronger result, if $F + \Phi$ are strictly convex.

Corollary 1. *If the set of stationary points consists of only one point, then $\{u_n\}$ converges to the unique solution of the minimization problem*

$$\min_{u \in H} \; F(u) + \Phi(u) \; .$$

As mentioned above, this situation occurs e. g. if $F+\Phi$ is strictly convex. Once more, the functional F itself does not need to be convex, thus the corollary also states the convergence of generalized conditional gradient methods associated with such functionals.

Remark 3. Some conditions for the convergence theorem can be relaxed.

- A lower semi-continuous Φ is also weakly lower semi-continuous. So, the existence of a minimizer also follows if the E_t are weakly compact (which is the case when Φ fulfills Condition 1) and if F is bounded and weakly lower semi-continuous.
- The weak lower semi-continuity of Ψ can also be established, if F' is completely continuous.

4 Connections between the Methods

The generalized conditional gradient method as well as the iterative shrinkage method based on surrogate functionals have been introduced in Section 2. In this section we show that the iterative shrinkage method can be interpreted as a generalized conditional gradient method. Furthermore, we analyze the application of the generalized conditional gradient method to inverse problems.

4.1 The Iterative Shrinkage Method as a Generalized Conditional Gradient Method

The surrogate approach and the related iteration scheme have been introduced for solving inverse problems with non-standard penalty terms (4):

$$\min_{u\in H} \frac{\|Ku-f\|^2}{2} + \sum_n w_n |\langle u,\, \varphi_n\rangle|^p \quad .$$

We will now apply the generalized conditional gradient method to this particular problem. We will see, that this reproduces the iteration scheme proposed in [3] for solving (4). For that purpose, we split the functional in (4) such that the minimization problem can be written as

$$\min_{u\in H}\ F(u) + \Phi(u)$$

with a $\lambda > 0$,

$$F(u) = \frac{\|Ku-f\|^2 - \lambda\|u\|^2}{2} \quad , \qquad \Phi(u) = \frac{\lambda\|u\|^2}{2} + \sum_n w_n |\langle u,\, \varphi_n\rangle|^p \ . \tag{9}$$

Note that the derivative of F is given by

$$F'(u) = K^*(Ku-f) - \lambda u \quad .$$

We now adapt the steps of the generalized conditional gradient method for this special type of functionals. First of all we need to determine a descent direction v. The linearized problem for determining a descent direction for the generalized conditional gradient method reads as

$$\min_{v\in H} \langle K^*(Ku-f) - \lambda u,\, v\rangle + \frac{\lambda\|v\|^2}{2} + \sum_n w_n |\langle v,\, \varphi_n\rangle|^p \quad .$$

This is equivalent to

$$\min_{v\in H} \sum_n \tfrac{1}{2} |\langle K^*(Ku-f) - \lambda u + \lambda v,\, \varphi_n\rangle|^2 + \lambda w_n |\langle v,\, \varphi_n\rangle|^p \ . \tag{10}$$

Except for the presence of λ, this is exactly the same problem as (5). In analogy to (6) its solution is given by

$$v = \mathbf{S}_{w/\lambda,p}\big(u - \lambda^{-1} K^*(Ku-f)\big) \tag{11}$$

with the shrinkage operator $\mathbf{S}_{w,p}$ defined as in Section 2.

The next step of the generalized conditional gradient method requires to determine a step size s. This reduces to a one-dimensional minimization problem for the functional $F + \Phi$. It amounts to

$$\min_{s\in[0,1]} \frac{\|K(u+s(v-u)) - f\|^2}{2} + \sum_n w_n |\langle u + s(v-u),\, \varphi_n\rangle|^p$$

or equivalently

$$\begin{gathered}\min_{s\in[0,1]} \frac{(s-r)^2}{2} + q\sum_n w_n|a_n + b_n s|^p \quad , \\ q = \frac{1}{\|K(u-v)\|^2} \qquad , \quad r = \frac{\langle Ku - f,\, K(u-v)\rangle}{\|K(u-v)\|^2} \quad , \\ a_n = \langle u,\, \varphi_n\rangle \qquad , \quad b_n = \langle v - u,\, \varphi_n\rangle \ .\end{gathered} \tag{12}$$

The solution of this problem cannot be expressed explicitly for general $p \in \,]1,2[$ and has to be discussed separately. However, the special cases $p = 1$ and $p = 2$ can be solved explicitly. We will present a procedure for this task in the appendix.

Note that in contrast to other line-search problems (e. g. projected gradient line-search) only two evaluations of the operator K are necessary. This can be useful if K has a complicated structure such as the solution operator of a partial differential equation. Alternatively, one can choose an approximate step size which decreases the functional nevertheless. This is always the case as we have seen in Lemma 2.

The method based on surrogate functionals just skips the line-search and chooses the step size $s = 1$ in every iteration. The surrogate functional approach also chooses $\lambda = 1$. However, λ can be interpreted as a correction factor if $\|K\| < 1$ is not satisfied.

Combining this choice of a descent direction with $s = 1$ and $\lambda = 1$ shows that the iteration process of the generalized conditional gradient method generates the same iterates as (11). Consequently, we see that the iterative shrinkage method is a special case of the generalized conditional gradient method with fixed step size.

Theorem 2. *The iterative shrinkage method for the minimization of the functional*

$$\frac{\|Ku - f\|^2}{2} + \sum_n w_n|\langle u,\, \varphi_n\rangle|^p$$

is the same as the generalized conditional gradient method with

$$F(u) = \frac{\|Ku - f\|^2 - \|u\|^2}{2} \quad \text{and} \quad \Phi(u) = \frac{\|u\|^2}{2} + \sum_n w_n|\langle u,\, \varphi_n\rangle|^p$$

and step size $s = 1$.

4.2 Application to Minimization in Linear and Non-Linear Inverse Problems

In the last section we have seen, that the generalized conditional gradient method yields the same iterates as the surrogate approach when applied to the regularized linear inverse problem

$$\min_{u\in H} \frac{\|Ku - f\|^2}{2} + \sum_n w_n|\langle u,\, \varphi_n\rangle|^p \quad .$$

In order to apply the convergence result in Theorem 1 we need to check the necessary assumptions on F and Φ. Note that this particular optimization problem is convex.

The application of the generalized conditional gradient method to (4) is based on the functionals F and Φ as defined in (9). One easily sees, that F always meets the requirements of Theorem 1. However, the assumption on Φ, that the sub-level sets $\{u \in H \ : \ \Phi(u) \leq t\}$ are compact, is crucial. In our special setting, this requirement requests that

$$E_t = \Big\{ u \in H \ : \ \sum_n w_n |\langle u, \varphi_n \rangle|^p \leq t \Big\}$$

is compact. In order to analyze the compactness of the sets E_t, we introduce the following spaces:

$$B_w^p = \{u : \mathbf{N} \to \mathbf{R} \ : \ \|u\|_{w,p} < \infty\} \ , \qquad \|u\|_{w,p} = \Big(\sum_n w_n |u_n|^p \Big)^{1/p}$$

Since the Hilbert space H is assumed to be separable and $\{\varphi_n\}$ is an orthonormal basis, the compactness criterion is fulfilled if and only if B_w^p is compactly embedded in ℓ^2. A sufficient condition for this, which can be easily proven by finite-dimensional approximation, is stated in the following lemma.

Lemma 6. *If $w_n \to \infty$ for $n \to \infty$, then B_w^p is compact in ℓ^2 for every $1 \leq p \leq 2$.*

Proof. Let us approximate the identity operator $I : B_w^p \to \ell^2$ by cutting off after the first k coefficients, i. e.

$$P_k(u)_n = \begin{cases} u_n & n \leq k \\ 0 & n > k \end{cases} .$$

Since all P_k are compact, we only have to show that I is the limit of P_k for $k \to \infty$ in operator norm. For a u with $\|u\|_{w,p} \leq 1$ there follows $w_n |u_n|^p \leq 1$ and since the conjugated exponent $p' \leq p$, there is $w_n^{p'/p} |u_n|^{p'} \leq w_n |u_n|^p$.

This helps to estimate the norm of $\|(I - P_k)u\|_2$ by

$$\begin{aligned}
\|(I - P_k)u\|_2 &= \Big(\sum_{n=k+1}^{\infty} w_n^{1/p} |u_n| w_n^{-1/p} |u_n| \Big)^{1/2} \\
&\leq \Big(\sum_{n=k+1}^{\infty} w_n |u_n|^p \Big)^{1/2p} \Big(\sum_{n=k+1}^{\infty} w_n^{-p'/p} |u_n|^{p'} \Big)^{1/2p'} \\
&\leq \Big(\sum_{n=k+1}^{\infty} w_n |u_n|^p \Big)^{1/2p} \Big(\sum_{n=k+1}^{\infty} w_n^{-2p'/p} w_n |u_n|^p \Big)^{1/2p'} \\
&\leq \max_{n \geq k+1} w_n^{-1/p} \Big(\sum_{n=k+1}^{\infty} w_n |u_n|^p \Big)^{1/2} \leq \max_{n \geq k+1} w_n^{-1/p} \ .
\end{aligned}$$

Since $w_k \to \infty$ for increasing k, the norm gets small uniformly in u, thus the P_k converge to I in strong operator sense. □

Hence, Theorem 1 is applicable if $\{w_n\}$ and p are chosen according to the above lemma. In this case, the convergence of a subsequence of the iterates to a stationary point is established. But, since we know that the original problem (4) is convex, we can conclude that each stationary point is in fact a solution of (4). Moreover, if the functional in (4) is also strictly convex, then this solution has to be unique and the sequence moreover converges to this unique solution. For $p > 1$ the penalty term $\sum_n w_n |\langle u, \varphi_n \rangle|^p$ possesses this property. For $p = 1$ the uniqueness of the solution can be deduced if K is injective.

In this general form, the main result regarding the convergence of the generalized conditional gradient method for linear inverse problems can be summarized as follows.

Theorem 3. *Let problem* (4) *be given with* $w_n \to \infty$ *for* $n \to \infty$. *Assume* $1 < p \leq 2$ *or* $p = 1$ *and* K *an injective operator. Then the sequence generated by the generalized conditional gradient method when applied to* (9) *converges to the unique solution of problem* (4) *for every* $\lambda > 0$.

Remark 4. Regarding the generalized conditional gradient method and the iterative shrinkage method as equivalent, this theorem yields the convergence of the method of iterative shrinkage for the special case $w_n \to \infty$ as $n \to \infty$. In [3], the authors obtain convergence for the more general case $w_n \geq c > 0$ by exploiting the particular structure of problem (4). In contrast to that, the convergence result presented here can be derived directly from the general convergence theorem (Theorem 1), which also includes non-convex functionals as well as penalty terms not based on basis expansions.

Remark 5. We now consider the problem (4) where $K : H \to H$ is a continuously Fréchet differentiable, but not necessarily linear, operator. In this case the generalized conditional gradient algorithm can also be applied.

Given an iterate u, the descent direction v is then computed by

$$v = \mathbf{S}_{w/\lambda,p}\big(u - \lambda^{-1} K'(u)^*(K(u) - f)\big) \ .$$

Hence, it also resembles the iterative shrinkage method for non-linear K. In this case, the convergence theorem is applicable and one gets (at least) a subsequence which converges to a stationary point of the corresponding minimization problem. Of course, algorithms based on first order necessary conditions do not "see" whether such a point is a local or global minimum, thus one cannot expect more without further assumptions on K.

An independent investigation of a similar problem is considered in [14, 15]. The approach taken there is based on surrogate functionals. The obtained algorithm consists of two nested iterations and possesses similar convergence properties. These works also cover the case $w_n \geq c > 0$ under additional requirements on the operator K.

5 Summary

The starting point for this paper are constrained optimization problems. A natural generalization leads to problems

$$\min_{u \in H} \ F(u) + \Phi(u) \quad ,$$

where F is smooth, but not necessarily convex, and Φ is convex, but not necessarily smooth. Hence, the sum of these two functionals may be neither smooth nor convex. This generalization reduces to the standard constrained optimization problem if Φ is the indicator function of the admissible set.

The classical gradient method for constrained optimization problems now motivates a generalized gradient method for solving this problem, see Section 2. The first main result of this paper proves the convergence of this generalized gradient method under some natural assumptions on F and Φ.

We also investigate the application of the generalized conditional gradient method for solving inverse problems, see Section 4. In particular we show, that the generalized conditional gradient method coincides with the surrogate approach for minimizing functionals of type

$$\min_{u \in H} \frac{\|Ku - f\|^2}{2} + \sum_n w_n |\langle u, \varphi_n \rangle|^p \quad .$$

The convergence results of the generalized conditional gradient method extends to non-linear inverse problems as well.

Acknowledgements

The work of Kristian Bredies has been supported by DFG SPP 1114-MA 1657/14-1, Dirk Lorenz has been supported by the European Union's Human Potential Programme under contract HPRN-CT-2002-00285 (HASSIP), and Peter Maass has been supported by DFG MA 1657/15-1. This is gratefully acknowledged.

A Appendix

A.1 Solving the line-search problem

In each step, the generalized conditional gradient method requires a line-search which involves solving a one-dimensional minimization problem. In case of the inverse problem (4), this can be written as the sum of a parabola and an infinite series depending on s, see (12). In general, this cannot be solved analytically. Nevertheless, there are some special cases (e. g. $p = 1$ or $p = 2$) which allow an explicit computation of the minimizing s.

In the following, let u and v be given such that $\sum_n w_n |\langle u, \varphi_n \rangle|^p < \infty$ and define the coefficients $q, r, \{a_n\}, \{b_n\}$ as in (12).

A.1.1 The Case $p = 2$

Here, the infinite sum in the one-dimensional minimization problem reduces to

$$\sum_n w_n (a_n + b_n s)^2 = \sum_n w_n a_n^2 + 2s \sum_n w_n a_n b_n + s^2 \sum_n w_n b_n^2 = as^2 + 2cs + b^2 \quad .$$

Note that the sums b and c exist since

$$\sum_n w_n(\langle u-v,\, \varphi_n\rangle)^2 \leq 2\sum_n w_n(\langle u,\, \varphi_n\rangle)^2 + 2\sum_n w_n(\langle v,\, \varphi_n\rangle)^2 \ .$$

The latter is finite since v is required to be the solution of the minimization problem (10). Now it is easy to see that the line-search problem is equivalent to

$$\min_{s\in[0,1]} \ \frac{1}{2}\Big(s - \frac{r-2qc}{1+2qa}\Big)^2$$

whose solution is the projection $P_{[0,1]}\big((r-2qc)/(1+2qa)\big)$. The formula for the step size is given accordingly by

$$s = P_{[0,1]}\left(\frac{\langle Ku-f,\, K(u-v)\rangle + 2\sum_n w_n\langle u,\, \varphi_n\rangle\langle u-v,\, \varphi_n\rangle}{\|K(u-v)\|^2 + 2\sum_n w_n(\langle u_n,\, \varphi_n\rangle)^2}\right) \ .$$

A.1.2 The Case $p = 1$

This case is a bit more complicated. We need to introduce subgradients which will be a convenient tool for this type of convex minimization problems.

Let the functional $\Phi : H \to \mathbf{R} \cup \{\infty\}$ be convex. An $w \in H$ is a subgradient of g at $u \in H$, if

$$\Phi(u) + \langle w,\, v-u\rangle \leq \Phi(v) \quad \text{for all } v \in H \ .$$

We denote this by $w \in \partial\Phi(u)$, the collection of all subgradients is denoted by $\partial\Phi$.

The subgradient coincides with the usual derivative for differentiable functions.

Example 2. Let $H = \mathbf{R}^d$ and $\Phi(u) = \sum_n w_n|u_n|$. Then the subdifferential $\partial\Phi$ is given by

$$v \in \partial\Phi(u) \quad \Leftrightarrow \quad \forall n : v_n \in w_n \operatorname{sgn}(u_n) \ , \qquad \operatorname{sgn}(s) = \begin{cases} -1 & s < 0 \\ [-1,1] & s = 0 \\ 1 & s > 0 \end{cases} \ .$$

The solution of a convex minimization problem can be equivalently expressed by using subgradients:

$$g(s) = \min_{t\in\mathbf{R}} \ g(t) \quad \Leftrightarrow \quad 0 \in \partial g(s) \ .$$

We will only need one basic rule of subgradient calculus: If $g = g_1 + g_2$ is proper, i. e. it has a finite value at some point, and if g_1 is continuous, then $\partial g = \partial g_1 + \partial g_2$ (cf. [7]). This can be exploited to derive an equivalent formulation of the line-search problem:

$$\begin{aligned} & s = \operatorname*{argmin}_{t\in[0,1]} \ \frac{(t-r)^2}{2} + q\sum_n w_n|a_n + b_n t| \\ \Leftrightarrow \quad & 0 \in s - r + \partial h(s) \qquad h(s) = q\sum_n w_n|a_n + b_n s| + I_{[0,1]}(s) \ . \end{aligned}$$

In order to calculate the subgradient of h, we suppose that there are only finitely many nonzero coefficients b_n. We denote the set of these coefficients by N. Thus one can interpret the sum as a continuous functional in a finite-dimensional Hilbert space and the chain rule yields

$$\partial h(s) = q \sum_{n \in N} w_n b_n \operatorname{sgn}(a_n + b_n s) + \begin{cases}]-\infty, 0] & s = 0 \\ 0 & s \in]0,1[\\ [0, \infty[& s = 1 \end{cases} .$$

We introducing some short-hand notation as follows

$$\partial h(s) = c(s) + [d_1(s), d_2(s)] \ ,$$

with $t_n = -a_n/b_n$ for $n \in N$ and

$$c(s) = q \sum_{s > t_n} w_n |b_n| - q \sum_{s < t_n} w_n |b_n|$$

$$d_1(s) = \begin{cases} -\infty & s = 0 \\ -q \sum_{s = t_n} w_n |b_n| & s > 0 \end{cases} \qquad d_2(s) = \begin{cases} q \sum_{s = t_n} w_n |b_n| & s < 1 \\ \infty & s = 1 \end{cases} .$$

The minimization problem for determining the step size s now reads as: For a given $r \in \mathbf{R}$, find an $s \in [0, 1]$ such that

$$r \in s + c(s) + [d_1(s), d_2(s)] \ .$$

Note that the subgradient has to be monotone with respect to s, so the right hand side can be interpreted as a function with slope one and positive jumps at t_n. The subgradient ∂h as well as $I + \partial h$ are illustrated in Figure 1. As can be seen in the right picture, it is intuitively easy to find an $r \in (I + \partial h)(s)$. One just has to sort the t_n, start with $s = 0$ and increase s until r is reached.

Thus, the following algorithm for the solution of the problem is proposed, which amounts to a formalization of the above approach.

1. Set $k = 0$, $t_0 = 0$ and $d_0^1 = -\infty$, $d_0^2 = c(0) + d_2(0)$.
2. Find the smallest t_{k+1} satisfying $t_{k+1} > t_k$. If it does not exist or it is greater than 1, then let $t_{k+1} = 1$.
3. Let $d_{k+1}^1 = d_k^2 + (t_{k+1} - t_k)$.
4.
   - If $d_k^1 \le r \le d_k^2$, then the we found the solution $s = t_k$, thus quit.
   - If $d_k^2 \le r \le d_{k+1}^1$, then the solution is $s = r - d_k^2$, thus quit.
5. If $t_{k+1} = 1$, then the solution is $s = 1$, thus quit. Else set $k = k + 1$, let $d_k^2 = d_k^1 + 2q \sum_{t_n = t_k} w_n |b_n|$ and continue with Step 2.

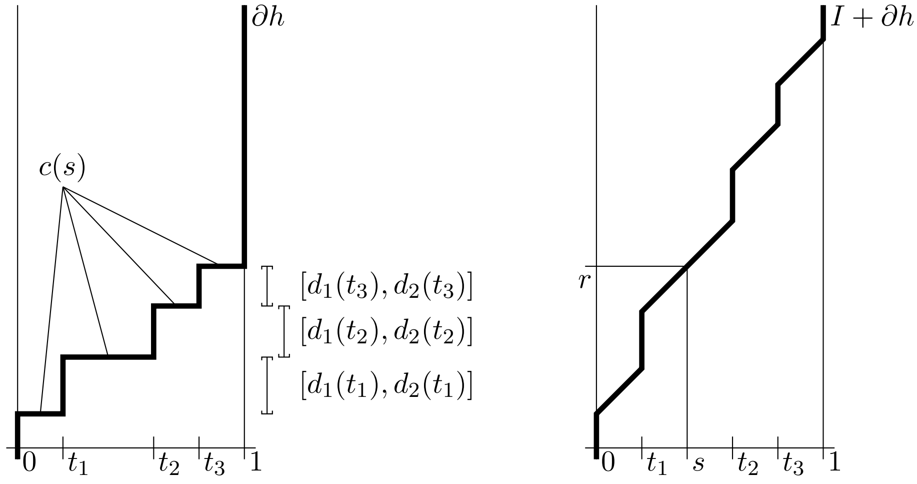

Figure 1: The line-search minimization problem in terms of subgradients. On the left hand side, the structure of ∂h as a monotone "staircase" is illustrated while on the right hand side, it is shown how the sought minimization point s is related to the given r.

We initially supposed that there are only finitely many nonzero coefficients b_n. This guarantees, that there are only finitely many t_n which guarantees that the algorithm will terminate.

We will finish this section by analyzing the assumption, that only finitely many coefficients b_n are non-zero. The $\{b_n\}$ are the coefficients of the difference $v - u$ with respect to the basis $\{\varphi_n\}$. It is therefore sufficient, that both u and v are a linear combination of finitely many φ_n. If the weights satisfy $w_n \geq w > 0$, then one can see from (11) that this is always fulfilled for v. Hence, if the n-th iterate u_n has a finite representation with respect to the basis, then the $(n+1)$-th iterate will share this property. Thus choosing for instance $u_0 = 0$ ensures the applicability of the above algorithm for each iteration of the generalized conditional gradient method.

References

[1] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16:1–3, 1966.

[2] H. Brezis. *Opérateurs maximaux monotones et semi-groups de contractions dans les espaces de Hilbert*, volume 5 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., 1973.

[3] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[4] V. F. Demyanov and A. M. Rubinov. *Approximate Methods in Optimization Problems.* Number 32 in Modern Analytic and Computational Methods in Science and Mathematics. American Elsevier, 1970.

[5] J. C. Dunn. Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals. *SIAM Journal on Control and Optimization*, 17(2):187–211, March 1979.

[6] J. C. Dunn. Convergence rates for conditional gradient sequences generated by implicit step length rules. *SIAM Journal on Control and Optimization*, 18(5):473–487, September 1980.

[7] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North-Holland, Amsterdam, 1976.

[8] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 2000.

[9] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quaterly*, 3:95–110, 1956.

[10] A. A. Goldstein. On steepest descent. *SIAM Journal on Control and Optimization*, 3:147–151, 1965.

[11] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer, 1971.

[12] A. K. Louis. *Inverse und schlecht gestellte Probleme.* B.G. Teubner, 1989.

[13] E. Polak. An historical survey of computational methods in optimal control. *SIAM Review*, 15(2):553–584, 1973.

[14] R. Ramlau and G. Teschke. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Problems*, 21(5):1571–1592, 2005.

[15] R. Ramlau and G. Teschke. A Tikhonov based projection iteration for nonlinear ill posed problems with sparsity constraints. Accepted for publication in *Numerische Mathematik*, 2006.

[16] E. Zeidler. *Nonlinear Functional Analysis and its Applications III.* Springer-Verlag, 1984.

Kristian Bredies, Peter Maass
Fachbereich 3
Universität Bremen
Postfach 33 04 40
28334 Bremen

Germany

Dirk A. Lorenz
Electrical Engineering Department
Technion - Israel Institute of Technology
32000 Haifa
Israel