On the constrained minimization of smooth Kurdyka—Łojasiewicz functions with the scaled gradient projection method

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2016 J. Phys.: Conf. Ser. 756 012001
(http://iopscience.iop.org/1742-6596/756/1/012001)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 155.185.102.48
This content was downloaded on 02/11/2016 at 15:13

Please note that terms and conditions apply.

You may also be interested in:

An image reconstruction method from Fourier data with uncertainties on the spatial frequencies
Anastasia Cornelio, Silvia Bonettini and Marco Prato

Stochastic variational approach to minimum uncertainty states
F Illuminati and L Viola

Interactions between polyelectrolyte-macroion complexes
H. Boroudjerdi and R. R. Netz

Design of copolymeric materials
T Kurosky and J M Deutsch

On the linear statistics of Hermitian random matrices
Yang Chen and Nigel Lawrence

Theoretical description of the nucleation of vapor bubbles in a superheated fluid
J. F. Lutsko

Numerical identifications of parameters in parabolic systems
Yee Lo Keung and Jun Zou
On the constrained minimization of smooth Kurdyka–Łojasiewicz functions with the scaled gradient projection method

Marco Prato¹, Silvia Bonettini², Ignace Loris³, Federica Porta² and Simone Rebegoldi¹

¹ Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy
² Dipartimento di Matematica e Informatica, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
³ Département de Mathématique, Université Libre de Bruxelles, Boulevard du Triomphe, 1050 Bruxelles, Belgium
E-mail: marco.prato@unimore.it

Abstract. The scaled gradient projection (SGP) method is a first-order optimization method applicable to the constrained minimization of smooth functions and exploiting a scaling matrix multiplying the gradient and a variable steplength parameter to improve the convergence of the scheme. For a general nonconvex function, the limit points of the sequence generated by SGP have been proved to be stationary, while in the convex case and with some restrictions on the choice of the scaling matrix the sequence itself converges to a constrained minimum point. In this paper we extend these convergence results by showing that the SGP sequence converges to a limit point provided that the objective function satisfies the Kurdyka–Łojasiewicz property at each point of its domain and its gradient is Lipschitz continuous.

1. The scaled gradient projection method
The scaled gradient projection algorithm (SGP) [4] belongs to the class of first-order methods designed to solve any constrained optimization problem of the form

\[ \min_{x \in \mathbb{R}^n} \Psi(x) \equiv f(x) + \iota_\Omega(x), \]  

(1)

where \( f \) is a continuously differentiable function and \( \iota_\Omega \) is the indicator function of the nonempty, closed and convex set \( \Omega \). In particular, the \((k+1)\)-th SGP iteration is computed as

\[ x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} = x^{(k)} + \lambda_k \left( P_{\Omega, D_k^{-1}} (x^{(k)} - \alpha_k D_k \nabla f(x^{(k)})) - x^{(k)} \right), \]

where \( \alpha_k \) is a scalar steplength parameter, \( D_k \) is a symmetric positive definite matrix and \( P_{\Omega, D_k^{-1}}(\cdot) \) is the projection onto \( \Omega \) associated to the norm induced by \( D_k^{-1} \) [4]. The step along
2. Preliminary results

Let \( \Psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) be a proper, lower semicontinuous function. For \( -\infty < \eta_1 < \eta_2 \leq +\infty \), let us set \( [\eta_1, \eta_2] = \{ z \in \mathbb{R}^n : \eta_1 < \Psi(z) < \eta_2 \} \). Moreover, we denote with \( \partial \Psi(z) \) the subdifferential of \( \Psi \) at \( z \in \mathbb{R}^n \) and with \( \text{dist}(z, \Omega) \) the distance between a point \( z \) and a set \( \Omega \subseteq \mathbb{R}^n \). The function \( \Psi \) is said to have the KL property at \( z \in \text{dom}\Psi := \{ z \in \mathbb{R}^n : \partial \Psi(z) \neq \emptyset \} \) if there exist \( \eta \in (0, +\infty) \), a neighborhood \( U \) of \( z \) and a continuous concave function \( \varphi : [0, \eta] \rightarrow [0, +\infty) \) such that \( \varphi(0) = 0 \), \( \varphi \) is \( C^1 \) on \( (0, \eta) \), \( \varphi'(s) > 0 \) for all \( s \in (0, \eta) \) and the KL inequality

\[
\varphi'(\Psi(z) - \Psi(\bar{z})) \text{dist}(0, \partial \Psi(z)) \geq 1.
\] (4)

holds for all \( z \in U \cap [\Psi(\bar{z}) < \Psi < \Psi(\bar{z}) + \eta] \). If \( \Psi \) satisfies the KL property at each point of \( \text{dom}\partial \Psi \), then \( \Psi \) is called a KL function.

The convergence proof of SGP for KL functions follows the ideas presented in [2], in which the authors proved an abstract convergence result for descent methods satisfying the following three conditions:

1. The descent direction \( d(k) \) is performed by means of the linesearch parameter \( \lambda_k = \delta^{m_k} \), where \( \delta \in (0, 1) \) and \( m_k \) is the smallest non-negative integer such that the monotone Armijo condition

\[
f(x(k) + \lambda_k d(k)) \leq f(x(k)) + \beta \lambda_k \nabla f(x(k))^T d(k)
\] (2)

is satisfied for a fixed value of the parameter \( \beta \in (0, 1) \).

2. The SGP method is a variable metric forward-backward algorithm [11, 12] which has been exploited in the last years for the solution of different real-world inverse problems [5, 6, 7, 16, 17, 18]. The main difference between SGP and the standard forward-backward schemes is the presence of two independent parameters \( \alpha_k \) and \( \lambda_k \) with a complete different role: while the last one is automatically computed with the Armijo condition (2) to guarantee the sufficient decrease of the objective function, the first one can be chosen to improve the actual convergence rate of the method, exploiting thirty years of literature in numerical optimization [3, 13, 19, 20].

3. We also remark that, unlike the schemes presented e.g. in [2, Sect. 5], in SGP the explicit computation of the steplength \( \alpha_k \) is forced to assume lower values than it originally would with SGP. By doing so, the new sequence satisfies the condition

\[
x(k+1) = \begin{cases} y(k) & \text{if } f(y(k)) < f(x(k) + \lambda_k d(k)) \\ x(k) + \lambda_k d(k) & \text{otherwise.}
\end{cases}
\] (3)

so that the sequence \( \{f(x(k))\}_{k \in \mathbb{N}} \) is forced to assume lower values than it originally would with SGP. By doing so, the new sequence satisfies the condition \( f(x(k+1)) \leq f(x(k) + \lambda_k d(k)) \), which still guarantees the stationarity of the limit points [8]. The aim of this paper is to prove the convergence of this modified SGP scheme if the (nonconvex) objective function \( \Psi \) satisfies the Kurdyka–Łojasiewicz (KL) property [15, 14], which holds true for most of the functions commonly used in inverse problems as \( p \) norms, Kullback-Leibler divergence and indicator functions of box plus equality constraints. A generalization of the proposed method and the related convergence proof to the minimization of the sum of smooth (nonconvex) and convex (nonsmooth) KL functions is in progress [10].
Lemma 1 The sequence \( \{x^{(k)}\}_{k \in \mathbb{N}} \) satisfies
\[
\frac{1}{2\mu \alpha_{\text{max}}} \| y^{(k)} - x^{(k)} \|^2 \leq -\nabla f(x^{(k)})^T (y^{(k)} - x^{(k)}).
\] (5)
Moreover, if \( \nabla f \) is \( L \)-Lipschitz continuous on \( \Omega \), then:
- there exists \( c > 0 \) such that for all \( \lambda \in [0, 1] \)
  \[
f(x^{(k)} + \lambda d^{(k)}) \leq f(x^{(k)}) + \lambda(1 - cL\lambda) \nabla f(x^{(k)})^T d^{(k)};
\] (6)
- the sequence \( \{\lambda_k\}_{k \in \mathbb{N}} \) of \( \text{SGP linesearch parameters} \) is bounded away from zero, i.e.,
  \[
  \lambda_k \geq \lambda_{\text{min}} \quad \forall k \in \mathbb{N}, \quad \lambda_{\text{min}} > 0.
\] (7)
- if \( f \) is also bounded from below, then \( 0 < - \sum_{k=0}^{+\infty} \nabla f(x^{(k)})^T d^{(k)} < +\infty \) and, consequently,
  \[
  \lim_{k \to +\infty} \nabla f(x^{(k)})^T d^{(k)} = 0.
\] (8)

Corollary 1 If \( \nabla f \) is \( L \)-Lipschitz continuous on \( \Omega \), then
\[
\Psi(x^{(k+1)}) + a \| x^{(k+1)} - x^{(k)} \|^2 \leq \Psi(x^{(k)}).
\] (H1)

Proof: Combining (5) with the backtracking rule (2) immediately yields
\[
f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{\beta \lambda_k}{2\mu \alpha_{\text{max}}} \| y^{(k)} - x^{(k)} \|^2.
\] (9)
Because of (3), it is either \( x^{(k+1)} = y^{(k)} \) or \( x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} \). In the first case, using (9) with (7) leads to
\[
f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{\beta \lambda_{\text{min}}}{2\mu \alpha_{\text{max}}} \| x^{(k+1)} - x^{(k)} \|^2.
\] (10)
In the second case, we obtain the same inequality by using \( y^{(k)} - x^{(k)} = \frac{1}{\lambda_k} (x^{(k+1)} - x^{(k)}) \) and \( \lambda_{\text{min}} \leq \lambda_k \leq 1 \) in (9). Finally, (H1) follows by adding the indicator function \( \iota_{\Omega} \) to both terms of (10) and by taking \( a = \frac{\beta \lambda_{\text{min}}}{2\mu \alpha_{\text{max}} \beta} \).

We are now able to prove a slight variant of condition H2 for \( \text{SGP} \), in which a subgradient of \( \Psi \) in \( y^{(k)} \) (and not in \( x^{(k)} \)) is provided.

Lemma 2 Suppose that \( \nabla f \) is \( L \)-Lipschitz continuous on \( \Omega \). There exist \( b > 0 \) and \( v^{(k)} \in \partial \Psi(y^{(k)}) \) for all \( k \in \mathbb{N} \) such that
\[
\| v^{(k)} \| \leq b \| x^{(k+1)} - x^{(k)} \|.
\] (H2)
Proof: By setting \( \tilde{h}_k(y) = \nabla f(x^{(k)})^T(y - x^{(k)}) + \frac{1}{\alpha_k} \|y - x^{(k)}\|^2 + \iota_\Omega(y) \), we rewrite \( y^{(k)} \) as
\[
y^{(k)} = \arg \min_{y \in \mathbb{R}^n} \tilde{h}_k(y) \iff 0 \in \partial \tilde{h}_k(y^{(k)}) \iff 0 \in \nabla f(x^{(k)}) + \frac{1}{\alpha_k} D_{k}^{-1}(y^{(k)} - x^{(k)}) + \partial \iota_\Omega(y^{(k)})
\]
\[
\iff \varpi^{(k)} = -\nabla f(x^{(k)}) + \frac{1}{\alpha_k} D_{k}^{-1}(x^{(k)} - y^{(k)}) \in \partial \iota_\Omega(y^{(k)}).
\]
Let us define \( v^{(k)} = \nabla f(y^{(k)}) + \varpi^{(k)} \in \partial \Psi(y^{(k)}) \) for all \( k \in \mathbb{N} \). By using the Lipschitz continuity of \( \nabla f \) and (7), the following inequalities hold true:
\[
\|v^{(k)}\| \leq \|\nabla f(y^{(k)}) - \nabla f(x^{(k)})\| + \frac{\mu}{\alpha_k} \|y^{(k)} - x^{(k)}\| \leq (L + \frac{\mu}{\alpha_{\min}}) \|y^{(k)} - x^{(k)}\|
\]
\[
\leq \frac{1}{\lambda_k} (L + \frac{\mu}{\alpha_{\min}}) \|x^{(k+1)} - x^{(k)}\| \leq \frac{1}{\lambda_{\min}} (L + \frac{\mu}{\alpha_{\min}}) \|x^{(k+1)} - x^{(k)}\|.
\]
The thesis holds by choosing \( b = \frac{1}{\lambda_{\min}} (L + \frac{\mu}{\alpha_{\min}}) \). \( \square \)

3. Convergence result
In this section we present the convergence proof of SGP, which differs from that in [2, Lemma 2.6] by the presence of the projection step at each \( k \in \mathbb{N} \), which was not considered in [2].

Theorem 1 Suppose \( \Psi \) is a KL function and \( \nabla f \) is \( L \)-Lipschitz continuous on \( \Omega \). If \( \bar{x} \) is a limit point of \( \{x^{(k)}\}_{k \in \mathbb{N}} \), then \( x \) is a stationary point for (1) and \( x^{(k)} \) converges to \( \bar{x} \).

Proof: The stationarity of \( \bar{x} \) has been proved in [4]. Since \( \Psi \) is a KL function, it satisfies the KL property at each point of \( \Omega \) and, in particular, at \( \bar{x} \). This means that there exist \( \eta, \varphi \) and \( U \) as in Section 2, such that the KL inequality (4) holds at \( \bar{x} \). Equation (H1) implies that \( \{\Psi(x^{(k)})\}_{k \in \mathbb{N}} \) is a non increasing sequence and hence \( \Psi(x^{(k)}) \to \gamma, \Psi(x^{(k)}) \geq \gamma \). By the continuity of \( \Psi \) on \( \Omega \) and the fact that \( \bar{x} \) is a limit point of \( \{x^{(k)}\}_{k \in \mathbb{N}} \), we deduce that \( \gamma = \Psi(\bar{x}) \). From (6) with \( \lambda = 1 \) and by definition (3) of \( x^{(k+1)} \) we obtain
\[
\Psi(x^{(k+1)}) \leq \Psi(y^{(k)}) \leq \Psi(x^{(k)}) + (1 - cL)\nabla f(x^{(k)})^T d^{(k)}.
\]
Since \( \Psi(x^{(k)}) \to \Psi(\bar{x}) \) and (8) holds true, from (11) we also have that \( \Psi(y^{(k)}) \to \Psi(\bar{x}) \).

Consequently, for all sufficiently large \( k \) we have
\[
\Psi(\bar{x}) \leq \Psi(x^{(k)}) \leq \Psi(y^{(k-1)}) < \Psi(\bar{x}) + \eta.
\]

Furthermore, let \( \rho > 0 \) be such that \( B(\bar{x}, \rho) \subset U \). Then using the continuity of \( \varphi \), one can choose \( k_0 \in \mathbb{N} \) such that both (12) and the following technical condition are satisfied:
\[
\|\bar{x} - x^{(k_0)}\| = 3 \sqrt{\frac{\Psi(x^{(k_0)}) - \Psi(\bar{x})}{a\lambda_{\min}^2}} + \frac{b}{a} \varphi(x^{(k_0)}) - \Psi(\bar{x}) < \rho.
\]

We will now use \( \{x^{(k)}\}_{k \in \mathbb{N}} \) to denote the sequence \( \{x^{(k+k_0)}\}_{k \in \mathbb{N}} \). Let us rewrite (H1) as
\[
\|x^{(k+1)} - x^{(k)}\| \leq \frac{\sqrt{\Psi(x^{(k)}) - \Psi(x^{(k+1)})}}{a},
\]
which, by recalling \( x^{(k+1)} - x^{(k)} = \lambda_k(y^{(k)} - x^{(k)}) \) and (7), yields also
\[
\|y^{(k)} - x^{(k)}\| \leq \sqrt{\frac{\Psi(x^{(k)}) - \Psi(x^{(k+1)})}{a\lambda_{\min}^2}}.
\]
Fix $k \geq 1$. We state that if $x^{(k)}, y^{(k-1)} \in B(\bar{x}, \rho)$, then
\[
2\|x^{(k+1)} - x^{(k)}\| \leq \|x^{(k)} - x^{(k-1)}\| + \varphi_k,
\] (16)
where $\varphi_k = \frac{b}{2} [\varphi(\Psi(x^{(k)}) - \Psi(\bar{x})) - \varphi(\Psi(x^{(k+1)}) - \Psi(\bar{x}))]$. Observe that, because of (12), the quantity $\varphi(\Psi(x^{(k)}) - \Psi(\bar{x}))$ makes sense for all $k \in \mathbb{N}$, and thus $\varphi_k$ is well posed.

If $x^{(k+1)} = x^{(k)}$ inequality (16) holds trivially. Then we assume $x^{(k+1)} \neq x^{(k)}$. This assumption, combined with (14) and (12), guarantees that $x^{(k)}, y^{(k-1)} \in B(\bar{x}, \rho) \cap |\Psi(\bar{x}) < \Psi < \Psi(\bar{x}) + \eta|$ and therefore we can use the KL inequality in $x^{(k)}$ and $y^{(k-1)}$.

By exploiting the KL inequality at $k$ last equation, the triangle inequality, (20) with $b$ equal to [2], we focus on the $\varphi$-function. Since $\varphi$ is concave, its derivative is non increasing, thus $\varphi'(\Psi(x^{(k)}) - \Psi(\bar{x})) \geq \varphi'(\Psi(y^{(k-1)}) - \Psi(\bar{x}))$. This fact applied to inequality (17) leads to
\[
\varphi'(\Psi(x^{(k)}) - \Psi(\bar{x})) \geq \frac{1}{b\|x^{(k)} - x^{(k-1)}\|}.
\] (18)

Then following the same procedure of [2, Lemma 2.6], which uses the concavity of $\varphi$, (H1) and (18), we obtain $\|x^{(k+1)} - x^{(k)}\|^2 \leq \varphi_k\|x^{(k)} - x^{(k-1)}\|$ which, by applying the inequality $2\sqrt{uv} \leq u + v$, gives relation (16).

We are now going to show that for $j = 1, 2, \ldots$
\[
x^{(j)}, y^{(j-1)} \in B(\bar{x}, \rho),
\] (19)
\[
\sum_{i=1}^{j} \|x^{(i+1)} - x^{(i)}\| + \|x^{(j+1)} - x^{(j)}\| \leq \|x^{(1)} - x^{(0)}\| + \chi_j,
\] (20)
where $\chi_j = \frac{b}{a}\varphi(\Psi(x^{(1)}) - \Psi(\bar{x})) - \varphi(\Psi(x^{(j+1)}) - \Psi(\bar{x}))$.

In order to prove (19)–(20), let us reason by induction on $j$. Using the triangle inequality, (14) with $k = 0$, the monotonicity of $\left\{\Psi(x^{(k)})\right\}_{k \in \mathbb{N}}$ and (13) we have
\[
\|\bar{x} - x^{(1)}\| \leq \|\bar{x} - x^{(0)}\| + \|x^{(0)} - x^{(1)}\| \leq \|\bar{x} - x^{(0)}\| + \sqrt{\frac{\Psi(x^{(0)}) - \Psi(\bar{x})}{a}} < \rho,
\]
namely $x^{(1)} \in B(\bar{x}, \rho)$. Using (15) with $k = 0$ and applying the same arguments as before, we also have $y^{(0)} \in B(\bar{x}, \rho)$. Finally, direct use of (16) shows that (20) holds with $j = 1$.

By induction, suppose that (19)–(20) hold for some $j \geq 1$. Since proving that $x^{(j+1)} \in B(\bar{x}, \rho)$ is identical to [2], we focus on $y^{(j)} \in B(\bar{x}, \rho)$. We rewrite (15) by noticing that $\Psi(\bar{x}) \leq \Psi(x^{(k+1)}) \leq \Psi(x^{(k)}) \leq \Psi(x^{(0)})$, which implies that $\|y^{(j)} - x^{(j)}\| \leq \sqrt{\Psi(x^{(0)}) - \Psi(\bar{x})}/a\lambda_{\text{min}}^2$. By using this last equation, the triangle inequality, (20) with $k = j$ and (13), we have
\[
\|\bar{x} - y^{(j)}\| \leq \|\bar{x} - x^{(0)}\| + \|x^{(0)} - x^{(1)}\| + \sum_{i=1}^{j} \|x^{(i+1)} - x^{(i)}\| + \|x^{(j+1)} - x^{(j)}\| + \|x^{(j)} - y^{(j)}\|
\leq \|\bar{x} - x^{(0)}\| + 2\|x^{(0)} - x^{(1)}\| + \chi_j + \|x^{(j)} - y^{(j)}\|
\leq \|\bar{x} - x^{(0)}\| + 3\sqrt{\frac{\Psi(x^{(0)}) - \Psi(\bar{x})}{a\lambda_{\text{min}}^2}} + \frac{b}{a} \varphi(\Psi(x^{(0)}) - \Psi(\bar{x})) < \rho,
\]
From (20) it immediately follows $k = x$ or equivalently $\sum_{j=1}^{2^k} \| x^{(j+1)} - x^{(j)} \| + \varphi_{j+1}$. Adding the above inequality with (20) (with $k = j + 1$), which completes the induction proof.

From (20) it immediately follows $\sum_{j=1}^{2^k} \| x^{(j+1)} - x^{(j)} \| \leq \| x^{(1)} - x^{(0)} \| + \frac{b}{\rho} \sum_{i=1}^{+\infty} \| \Psi(x^{(i)}) - \Psi(\bar{x}) \|$ and therefore $\sum_{i=1}^{+\infty} \| x^{(i+1)} - x^{(i)} \| < +\infty$, which implies that the sequence $\{ x^{(k)} \}_{k \in \mathbb{N}}$ converges to some $x^*$. Since $\bar{x}$ is a limit point of the sequence, it must be $x^* = \bar{x}$. □

Acknowledgments
This work has been partially supported by MIUR under the two projects FIRB - Futuro in Ricerca 2012, contract RBFR12M3AC and PRIN 2012, contract 2012MTE38N, and by the Italian GNCS - INdAM. I. Loris is a Research Associate of the Fonds de la Recherche Scientifique.

References