An iterated $\ell_1$ Algorithm for Non-smooth Non-convex Optimization in Computer Vision
- Supplementary Material -

Peter Ochs\textsuperscript{1}, Alexey Dosovitskiy\textsuperscript{1}, Thomas Brox\textsuperscript{1}, and Thomas Pock\textsuperscript{2}

\textsuperscript{1} University of Freiburg, Germany
\{ochs,dosovits, brox\}@cs.uni-freiburg.de
\textsuperscript{2} Graz University of Technology, Austria
pock@icg.tugraz.at

Here we give a detailed proof of Lemma 1 of Subsection 3.2 (Convergence analysis). We also show that the functions $F(x) = F_1(x) + F_2(|x|)$ we optimize in applications fulfill the technical Conditions (C1) and (C2) which are necessary for the results of Subsection 3.2 to hold. We remind what these conditions are:

\begin{itemize}
  \item[(C1)] $F_2$ is twice continuously differentiable in $\mathcal{H}_+$ and there exists a subspace $\mathcal{H}_c \subset \mathcal{H}$ such that for all $x \in \mathcal{H}_+$ holds: $h^\top \partial^2 F_2(x) h < 0$ if $h \in \mathcal{H}_c$ and $h^\top \partial^2 F_2(x) h = 0$ if $h \in \mathcal{H}_c^\perp$.
  \item[(C2)] $F_2(|x|)$ is a $C^1$-perturbation of a convex function, i.e., can be represented as a sum of a convex function and a $C^1$-smooth function.
\end{itemize}

We start by proving Lemma 1:

\textbf{Lemma 1} (see Subsection 3.2 of the main text). \textit{Let $(x^k)$ be the sequence generated by the Algorithm (3) and suppose $(x^k)$ is bounded and the Condition (C1) holds for $F_2$. Then}

\begin{equation}
\lim_{k \to \infty} (\partial F_2(|x^k|) - \partial F_2(|x^{k+1}|)) = 0. \tag{1}
\end{equation}

\textit{Proof}. Taylor theorem for $F_2$ gives:

\begin{align*}
F_2(|x^k|) - F_2(|x^{k+1}|) &= (\Delta^k)^\top \partial F_2(|x^k|) \\
- &\frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k,
\end{align*}

where $\Delta^k := |x^k| - |x^{k+1}|$, $|\tilde{x}^k| \in [|x^k|, |x^{k+1}|]$. We use this to refine the inequalities (6) from the proof of Proposition 1 in Subsection 3.2 (see main text for details): \begin{align*}
F(x^k) - F(x^{k+1}) &= F_1(x^k) - F_1(x^{k+1}) + F_2(|x^k|) - F_2(|x^{k+1}|) \\
&\geq (d^{k+1})^\top (x^k - x^{k+1}) + (w^k)^\top \Delta^k \\
&- \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k = (A^k)^\top (x^k - x^{k+1}) \\
&+ (w^k)^\top ((|x^k| - |x^{k+1}|) - c^{-1}) \cdot (x^k - x^{k+1}) \\
&- \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k = (q^{k+1})^\top (A x^k - A x^{k+1}) \\
&+ (w^k)^\top (|x^k| - c^{-1} \cdot x^k) - \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k \\
&= (q^{k+1})^\top (b - b) + \sum_i w_i^k (|x_i^k| - c^{-1} x_i^k) \\
&- \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k \geq - \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k \geq 0.
\end{align*}

Therefore,

\begin{equation}
F(x^k) - F(x^{k+1}) \geq - \frac{1}{2} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k \geq 0,
\end{equation}

and, hence,

\begin{equation}
\lim_{k \to \infty} (\Delta^k)^\top \partial^2 F_2(|x^k|) \Delta^k = 0.
\end{equation}

Using definition of the space $\mathcal{H}_c$ we get that

\begin{equation}
\lim_{k \to \infty} (\Pr_{\mathcal{H}_c} \Delta^k)^\top \partial^2 F_2(|x^k|) \Pr_{\mathcal{H}_c} \Delta^k = 0, \tag{2}
\end{equation}

where $\Pr_{\mathcal{H}_c}$ denotes orthogonal projection onto $\mathcal{H}_c$. From boundedness of $(x^k)$ and negativity of $\partial^2 F_2 \big|_{\mathcal{H}_c}$ we conclude that there exists $\nu > 0$ such that for all $k$:

\begin{equation}
(\Pr_{\mathcal{H}_c} \Delta^k)^\top \partial^2 F_2(|x^k|) \Pr_{\mathcal{H}_c} \Delta^k \leq - \nu \|\Pr_{\mathcal{H}_c} \Delta^k\|^2.
\end{equation}

Together with (2) this gives

\begin{equation}
\lim_{k \to \infty} \|\Pr_{\mathcal{H}_c} \Delta^k\|^2 = 0. \tag{3}
\end{equation}

Now we note that

\begin{equation}
\partial F_2(|x^k|) - \partial F_2(|x^{k+1}|) = \partial^2 F_2(|x^k|) \Pr_{\mathcal{H}_c} \Delta^k,
\end{equation}

\begin{equation}
\lim_{k \to \infty} \|\Pr_{\mathcal{H}_c} \Delta^k\|^2 = 0.
\end{equation}
for some $|\hat{x}^k| \in [x^k; x^{k+1}]$. Together with (3) this completes the proof.

Now we show that the Conditions (C1) and (C2) actually hold for the functions $F_2$ used in applications, namely (cf. Equations (8) and (9) of the main text):

$$F_2(|x|) = \sum_i \lambda_i f(|x_i|),$$

where

$$f(|x_i|) = (|x_i| + \varepsilon)^p \text{ or } f(|x_i|) = \log(1 + \beta |x_i|),$$

with $\varepsilon > 0$, $\beta > 0$ and $\lambda_i \geq 0$, $\forall i$. Obviously, for both choices the functions are infinitely differentiable and concave in $\mathbb{R}_+$. Therefore it suffices to prove the following lemma:

**Lemma 3.** Let $F_2(|x|) = \sum_i \lambda_i f(|x_i|)$, where $\lambda_i \geq 0$, $\forall i$ and $f : \mathbb{R}_+ \to \mathbb{R}$ is increasing, twice continuously differentiable and has strictly negative second derivative. Then the Conditions (C1) and (C2) hold for $F_2$.

**Proof.** We start with proving (C1). Obviously,

$$\partial^2 F_2(x) = \text{diag}(\lambda_i f''(x_i))_{i=1}^{\dim(H)}, \text{ for } x \in H_+.$$  

Hence,

$$h^\top \partial^2 F_2(x) h = \sum_i \lambda_i f''(x_i) h_i^2.$$  

Denote by $\Lambda$ a diagonal operator with $\lambda_i$'s on diagonal. Then for $H_c = (\ker \Lambda)^\perp$ the desired condition holds.

Now we prove (C2). For each term of $F_2$ we perform the following decomposition:

$$\lambda_i f(|x_i|) = \lambda_i f'(0) |x_i| + \lambda_i (f(|x_i|) - f'(0)|x_i|).$$

The first summand is convex due to non-negativity of $\lambda_i$ and $f'(0)$. The second summand is continuously differentiable for $x_i \neq 0$ and its derivative equals

$$\begin{cases} 
\lambda_i (f'(x_i) - f'(0)), & \text{if } x_i > 0, \\
\lambda_i (-f'(-x_i) + f'(0)), & \text{if } x_i < 0.
\end{cases}$$

Both these values approach zero when $x_i$ approaches 0, so the function is differentiable at 0 and, therefore, continuously differentiable on $\mathbb{R}$.

We proved that each term of $F_2$ is a sum of a convex function and a $C^1$-smooth function. Sum of $C^1$-perturbations of convex functions is a $C^1$-perturbation of a convex function, so this completes the proof.  

$\square$