Iterated Lifting for Robust Cost Optimization

Christopher Zach
christopher.zach@crl.toshiba.co.uk

Guillaume Bourmaud
guillaume.bourmaud@u-bordeaux.fr

Toshiba Research Europe
Cambridge, UK

University of Bordeaux
Bordeaux, France

Abstract

Optimization of latent model parameters using robust formulations usually creates a large number of local minima due to the quasi-convex shape of the underlying robust kernel. Lifting the robust kernel, i.e. embedding the problem into a higher-dimensional space, leads to significantly better local minima in a range of 3D computer vision problems (e.g. [10, 11, 12]). In this work we propose to iterate this lifting construction to obtain a more gradual lifting scheme for a given target kernel. Thus, a robust kernel is not directly lifted against the (non-robust) quadratic kernel, but initially against a different, less robust kernel. This process is iterated until the quadratic kernel is reached to allow utilization of efficient non-linear least-squares solvers. We demonstrate in synthetic and real problem instances that iterated lifting generally reaches better local minima than IRLS and standard half-quadratic lifting.

1 Introduction

Many computer vision algorithms seek to estimate parameters from data containing outliers. Most of the time, the parameters are obtained by minimizing a cost function involving robust data terms that are able to mitigate the influence of outliers. In order to significantly reduce the influence of outliers it is common to use quasi-convex kernels such as Tukey's biweight or Geman-McClure (see Table 1). However the sum of theses quasi-convex kernels create (usually a large number of) local minima that may be harmful for the optimization algorithm. Consequently, it is important to use an optimization algorithm with the ability to escape bad local minima. One way to achieve this is by "half-quadratic lifting" the robust kernel, which consists in lifting the robust kernel w.r.t. a *quadratic basis kernel*, to obtain a higher-dimensional problem and to perform joint optimization w.r.t. both the parameters of interest and the lifting variables [11]. In doing so, the optimization algorithm is able to avoid early "classification" of data points as outliers and thus escapes bad local minima. In this paper, we propose to go beyond half-quadratic lifting by proposing *iterated lifting*, which essentially consists in lifting a robust kernel with more than one lifting variable and performing optimization in this even higher dimensional space (see Figure 1).

Related work   Lifting kernels, i.e the trick of expressing a robust kernel as the minimum of a family of quadratic kernels, dates back to the 1980s. It has gone through different names since its inception such as *line processes* [2, 5], *half-quadratic minimization* [4, 8] and more recently *switching constraints* [10]. The resulting lifted cost function was often minimized

(a) Target kernel $\phi_{\mathrm{Wel},1}(x)$ and two basis kernels $\phi_{\mathrm{Wel},\sqrt{2}}(x)$ and $\frac{x^2}{2}$

(b) $\phi_{\mathrm{Wel},1}(x)$ lifted w.r.t. a quadratic basis kernel $\frac{x^2}{2}$ (Half-quadratic lifting [10])

(c) $\phi_{\mathrm{Wel},1}(x)$ lifted w.r.t. a non-quadratic basis kernel $\phi_{\mathrm{Wel},\sqrt{2}}(x)$ (Proposed)
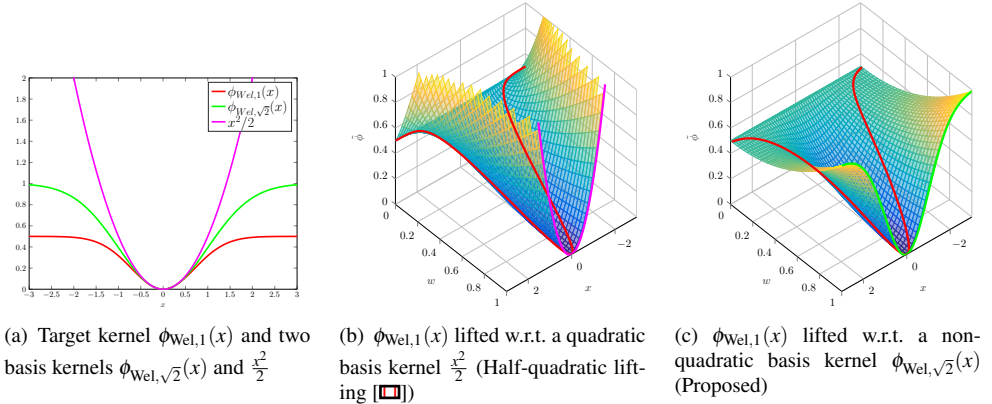
Figure 1: Comparison between half-quadratic lifting [10] and *iterated lifting* of the Welsch kernel (see Table 1). The *iterated lifting* technique proposed in this paper provides a gradual lifting scheme for a given target kernel as opposed to the standard Half-quadratic lifting technique. To do so, our approach lifts a target robust kernel w.r.t. a slightly less robust kernel. This process is iterated until the quadratic kernel is reached to allow utilization of efficient non-linear least-squares solvers.

using optimization by alternation, e.g Iterated Reweighted Least Squares (IRLS) [6]. However, this algorithm easily gets trapped into bad local minima.

It was only recently shown that optimizing in the higher dimensional parameter space of the lifted cost function directly, i.e jointly optimizing over the variables of interest as well as the lifting variables, could lead to better local minima [10]. This approach has been successfully employed in several applications since, e.g bundle adjustment [11], non-rigid reconstruction [12], monocular visual odometry [9], pose-graph optimization [10], scene reconstruction from RGB-D videos [3], multiview stereo [7].

Contributions   Our contributions are as follows:

1. We generalize the half-quadratic lifting construction to *non-quadratic basis kernels* and derive a technique to obtain the generalized lifted representation of a robust kernel.

2. We propose a novel technique, called *iterated lifting*, that allows to progressively lift a target robust kernel using a less robust kernel as basis kernel in an iterative manner.

3. We show that iteratively lifting a robust kernel w.r.t. a scaled version of itself is very convenient for at least three kernels of interest and demonstrate the effectiveness of our approach both on synthetic and real data.

Outline   The rest of the paper is organized as follows: Section 2 introduces some useful definitions and briefly reviews half-quadratic lifting. A generalized lifting formulation is presented in Section 3 which allows us to describe our iterated lifting framework in Section 4. In Section 5, our iterated lifting framework is evaluated experimentally. Finally, conclusions and future research directions are provided in Section 6.

2 Preliminaries

This section introduces some useful definitions and reviews the basis of half-quadratic lifting, which converts a difficult non-linear function into a higher-dimensional optimization problem that can be advantageous to solve. First we provide our definition of what we understand by a robust kernel:

Definition 1. A symmetric and sufficiently smooth function $\phi : \mathbb{R} \to \mathbb{R}_0^+$ is said to be a *robust kernel*, if

1. The mapping $z \mapsto \phi(\sqrt{2z})$ is monotonically increasing and concave in $\mathbb{R}_0^+$.

2. $\phi(0) = 0$, $\lim_{x \to 0} \phi'(x)/x = 1$ and $\phi''(0) = 1$,

The function $\omega(x) \stackrel{\text{def}}{=} \phi'(x)/x$ is called the *weight function* $\omega(x)$ of ϕ.

Note that we slightly deviate from the definition e.g. given in [4] by not limiting ourselves to redescending M-estimators and by normalizing ϕ differently. The first condition in Def. 1 is somewhat abstract, and at this point we aim to provide some intuition, since we will revisit and generalize this condition in Section 3.1. The mapping $\rho : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with $\rho(z) = \phi(\sqrt{2z})$ first converts the non-robust residual "energy" $z = z(x) = x^2/2$ to the corresponding residual magnitude $x(z) = \sqrt{2z}$ and then applies the robust kernel ϕ on residual x. Thus, ρ is a mapping between residual "energies", and the monotonicity and concavity of ρ means that this conversion between energies is diminishing for larger values (somewhat similar to monotone submodular functions).

Next we introduce a way to rewrite a robust kernel as a higher-dimensional problem:

Definition 2. Let ϕ be a robust kernel. If ϕ can be written as

$$\phi(x) = \min_{w \in [0,1]} \left\{ w \frac{x^2}{2} + \gamma(w) \right\} \stackrel{\text{def}}{=} \min_{w \in [0,1]} \tilde{\phi}(x, w) \tag{1}$$

for all x, then we call $\tilde{\phi}$ its *lifted representation* w.r.t. the *quadratic basis kernel* $x \mapsto x^2/2$.

In many cases γ can be continuously extended from $[0,1]$ to $\mathbb{R}_0^+$ (see Table 1), which is beneficial in numerical implementations since w can be reparametrized via $w = u^2$ or $w = \exp(u)$ instead of $w = \sigma(u)$ for a sigmoid function σ.

Finally we introduce a mapping $\zeta : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with $\zeta(x) = \sqrt{2\phi(x)}$ for a given robust kernel ϕ. ζ will be of interest in Section 4, and we state a simple fact:

Lemma 1. For a robust kernel ϕ the induced mapping ζ is a contraction, i.e. $\|\zeta(x)\| \le \|x\|$.

The lemma follows directly from $\phi(x) \le x^2/2$ since ϕ is a robust kernel.

3 Lifting using general basis kernels

3.1 Generalized lifting formulation

A natural question is whether we can replace the *quadratic basis kernel* $x \mapsto x^2/2$ in Eq. 1 by another basis kernel to obtain a generalized lifting formulation. The answer is positive, and will be the foundation of our iterated lifting construction.

Name	$\phi(x)$	$\omega(x)$	$\gamma(w)$	$\mathrm{dom}(\gamma)$
Quadratic	$x^2/2$	1	0	$\{1\}$
ℓ_1-ℓ_2	$\tau\sqrt{x^2+\tau^2}-\tau^2$	$\dfrac{\tau}{\sqrt{x^2+\tau^2}}$	$\frac{\tau^2}{2}(w+1/w)-\tau^2$	$\mathbb{R}_0^+$
Cauchy	$\frac{\tau^2}{2}\log\left(1+x^2/\tau^2\right)$	$\dfrac{\tau^2}{\tau^2+x^2}$	$\frac{\tau^2}{2}(w-\log w-1)$	$\mathbb{R}_0^+$
Huber	$\begin{cases} x^2/2 & \text{if } |x|\leq\tau \\ \tau|x|-\tau^2/2 & \text{if } |x|\geq\tau \end{cases}$	$\begin{cases} 1 & |x|\leq\tau \\ \tau/|x| & |x|\geq\tau \end{cases}$	$\frac{\tau^2}{2}(1/w-1)$	$[0,1]$
Geman-McClure	$\dfrac{\tau^2 x^2}{2(x^2+\tau^2)}$	$\dfrac{\tau^4}{(x^2+\tau^2)^2}$	$\frac{\tau^2}{2}(\sqrt{w}-1)^2$	$\mathbb{R}_0^+$
Welsch	$\frac{\tau^2}{2}\left(1-e^{-x^2/\tau^2}\right)$	e^{-x^2/τ^2}	$\frac{\tau^2}{2}(1+w\log w-w)$	$\mathbb{R}_0^+$
Truncated quadratic	$\min\{\tau,x\}^2/2$	$\begin{cases} 1 & |x|\leq\tau \\ 0 & |x|\geq\tau \end{cases}$	$\frac{\tau^2}{2}(1-w)$	$[0,1]$
Tukey's biweight	$\frac{\tau^2}{6}\left(1-\left[1-\frac{x^2}{\tau^2}\right]_+^3\right)$	$\left[1-x^2/\tau^2\right]_+^2$	$\frac{\tau^2}{6}(1-\sqrt{w})^2(1+2\sqrt{w})$	$\mathbb{R}_0^+$
Smooth truncated (ST)	$\frac{\tau^2}{4}\left(1-\left[1-\frac{x^2}{\tau^2}\right]_+^2\right)$	$\left[1-x^2/\tau^2\right]_+$	$\frac{\tau^2}{4}(w-1)^2$	$\mathbb{R}_0^+$

Table 1: Kernels (see Definition 1) and their related functions. All kernels are normalized such that $\phi''(0)=1$.

Proposition 1. Let ϕ and $\hat{\phi}:\mathbb{R}\to\mathbb{R}_0^+$ be robust kernels with $\hat{\phi}$ invertible in $\mathbb{R}_0^+$. $\phi\circ\hat{\phi}^{-1}$ is concave in $\mathbb{R}_0^+$ if and only if $\phi(x)=\min_w w\hat{\phi}(x)+\gamma(w)$ for some (convex) function γ.

Proof. We have the following property:

$$\phi(x)=\min_w w\hat{\phi}(x)+\gamma(w)=-\max_w\left\{-w\hat{\phi}(x)-\gamma(w)\right\}=-\gamma^*(-\hat{\phi}(x)). \tag{2}$$

where $\gamma^*(\cdot)$ is the convex conjugate of $\gamma(\cdot)$. As convex conjugate the function γ^* is convex in the argument $z(x)=-\hat{\phi}(x)$ (but not necessarily in x itself). This means that the hypograph of the mapping $-\hat{\phi}(x)\mapsto\phi(x)$ (with $x\geq0$) is convex. After changing coordinates $z=-\hat{\phi}(x)$ we obtain that the mapping $z\mapsto\phi(\hat{\phi}^{-1}(-z))$ is concave, which is equivalent to the concavity of the mapping $z\mapsto\phi(\hat{\phi}^{-1}(z))$.

If γ is chosen to be non-convex, then replacing γ by its bi-conjugate γ^{**} (which is convex by construction) in Eq. 2 does not alter the induced kernel ϕ (since $\gamma^{***}=\gamma^*=-\phi$). □

We consider the relation (also making use of $z=-\hat{\phi}(x)$),

$$-\phi(\hat{\phi}^{-1}(-z))=\gamma^*(z)=\max_w\left\{wz-\gamma(w)\right\}.$$

The properties of convex conjugates imply that the value of w attaining the maximum, $w^*(z)=\arg\max_w\{wz-\gamma(w)\}$, is always a (sub)gradient of γ^* at z. Assuming differentiability of γ^* (or ϕ) we therefore obtain

$$w^*(z)=\frac{d}{dz}\gamma^*(z)=-\frac{d}{dz}\phi(\hat{\phi}^{-1}(-z))=\frac{d}{dz}\phi(\hat{\phi}^{-1}(z)). \tag{3}$$

As shown in the next result there is a strong connection between w^* and the weight functions ω and $\hat{\omega}$ of the corresponding kernel ϕ and $\hat{\phi}$, respectively:

Proposition 2. Let ϕ and $\hat{\phi}$ be robust kernels. If $\phi\circ\hat{\phi}^{-1}$ is concave and non-decreasing in $\mathbb{R}_0^+$, then the following holds:

1. $\lim_{z\to0}\frac{d}{dz}\phi(\hat{\phi}^{-1}(z))=1$.

2. ϕ is below $\hat{\phi}$, i.e. $\phi(x) \leq \hat{\phi}(x)$.

3. The relative weight function $\omega/\hat{\omega}$ is non-negative and bounded by 1, and it is monotonically decreasing in $\mathbb{R}^+$.

Proof. First observe that (with $x = \hat{\phi}^{-1}(z)$)

$$\frac{d}{dz}\phi(\hat{\phi}^{-1}(z)) = \phi'(\hat{\phi}^{-1}(z))\frac{d}{dz}\hat{\phi}^{-1}(z) = \frac{\phi'(x)}{\hat{\phi}'(x)} = \frac{\omega(x)}{\hat{\omega}(x)}$$

$$\frac{d^2}{dz^2}\phi(\hat{\phi}^{-1}(z)) = \frac{\phi''(x)\hat{\phi}'(x) - \phi'(x)\hat{\phi}''(x)}{(\hat{\phi}'(x))^2}. \tag{4}$$

Hence, $\frac{d}{dz}\phi(\hat{\phi}^{-1}(z))|_{z=0} = \frac{\omega(0)}{\hat{\omega}(0)} = 1$, which shows the first claim. The concavity of $\phi \circ \hat{\phi}^{-1}$ implies that

$$\phi(\hat{\phi}^{-1}(z)) \leq \overbrace{\phi(\hat{\phi}^{-1}(0))}^{=0} + z \cdot \frac{d}{dz}\phi(\hat{\phi}^{-1}(z))|_{z=0} = z.$$

Choosing $z = \hat{\phi}(x)$ yields $\phi(x) \leq \hat{\phi}(x)$ as claimed. Finally, by recalling Eq. 4 we read

$$\frac{d}{dx}\frac{\omega(x)}{\hat{\omega}(x)} = \frac{\phi''(x)\hat{\phi}'(x) - \phi'(x)\hat{\phi}''(x)}{(\hat{\phi}'(x))^2} = \frac{d^2}{dz^2}\phi(\hat{\phi}^{-1}(z)),$$

which is non-positive by the concavity assumption. Thus, $\omega(x)/\hat{\omega}(x)$ is monotonically decreasing in $\mathbb{R}^+$. $\qquad\square$

Since $w^*(z) = \frac{d}{dz}\phi(\hat{\phi}^{-1}(z)) \in [0,1]$ (recall Eq. 3) we can add the constraint $w \in [0,1]$ in Eq. 2 w.l.o.g. This allows us to generalize the definition of lifted representations for robust kernels given earlier in Def. 2:

Definition 3. Let ϕ and $\hat{\phi} : \mathbb{R} \to \mathbb{R}_0^+$ be robust kernels such that $\hat{\phi}$ is invertible in $\mathbb{R}_0^+$ and $\phi \circ \hat{\phi}^{-1}$ is concave and non-decreasing in $\mathbb{R}_0^+$. Hence, $\phi(x)$ can be written as

$$\phi(x) = \min_{w \in [0,1]} \left\{ w\hat{\phi}(x) + \gamma(x) \right\}, \tag{5}$$

and we call the expression $w\hat{\phi}(x) + \gamma(x)$ the *lifted representation* of a robust kernel ϕ w.r.t. the basis kernel $\hat{\phi}$.

3.2 Obtaining Lifted Representations

Now that we know when ϕ can be lifted w.r.t. $\hat{\phi}$, the next question is how to obtain the expression of $\gamma(w)$. Let a pair of functions ϕ and $\hat{\phi}$ be given that satisfy the requirements in Prop. 2. A lifted representation of ϕ w.r.t. $\hat{\phi}$ can be obtained by using the ansatz

$$\phi(x) = \min_t \alpha(t)\hat{\phi}(x) + \beta(t), \tag{6}$$

and using the fact that function value and gradient need to match at $x = t$, i.e.

$$\phi(x)|_{x=t} = \alpha(t)\hat{\phi}(x)|_{x=t} + \beta(t) = \alpha(t)\hat{\phi}(t) + \beta(t)$$
$$\phi'(x)|_{x=t} = \alpha(t)\hat{\phi}'(x)|_{x=t} = \alpha(t)\hat{\phi}'(t). \tag{7}$$

We immediately obtain

$$\alpha(t) = \frac{\phi'(t)}{\hat\phi'(t)} = \frac{\omega(t)}{\hat\omega(t)} \qquad \beta(t) = \phi(t) - \alpha(t)\hat\phi(t) = \phi(t) - \frac{\omega(t)}{\hat\omega(t)}\hat\phi(t). \qquad (8)$$

By our assumptions we have $\alpha(t) \in [0,1]$ and $\beta(t) \geq 0$ for all t. If $\alpha(\cdot)$ is invertible in $[0,1]$, then reparametrizing $t = \alpha^{-1}(w)$ with $w \in [0,1]$ yields

$$\gamma(w) = \phi(\alpha^{-1}(w)) - \frac{\phi'(\alpha^{-1}(w))}{\hat\phi'(\alpha^{-1}(w))}\hat\phi(\alpha^{-1}(w)). \qquad (9)$$

Different reparametrizations of t are possible and could be beneficial for optimization, but in this work we use this canonical form of lifting. We found that this construction is more general and less error-prone than the criterion given in [11].

3.3 Lifting w.r.t. Scaled Robust Kernels

A natural way to lift a robust kernel ϕ is to use a scaled version of ϕ itself as basis kernel $\hat\phi$. Let ϕ be a robust kernel. For a scalar $\mu > 0$ we define the scaled kernel

$$\phi(x;\mu) = \mu^2\phi(x/\mu) \qquad (10)$$

(note that $\phi''(0) = 1$). We immediately obtain $\phi'(x;\mu) = \mu\phi'(x/\mu)$ and therefore

$$\omega(x;\mu) = \frac{\mu\phi'(x/\mu)}{x} = \frac{\phi'(x/\mu)}{x/\mu} = \omega(x/\mu). \qquad (11)$$

Below we will address the general case of lifting $\phi(\cdot;\lambda)$ w.r.t. $\phi(\cdot;\mu)$ for a $\lambda \in (0,\mu)$. Following the construction in Sec. 3.2 we read

$$\alpha(t) = \frac{\omega(t;\lambda)}{\omega(t;\mu)} = \frac{\omega(t/\lambda)}{\omega(t/\mu)} \qquad (12)$$

$$\beta(t) = \phi(t;\lambda) - \alpha(t)\phi(t;\mu) = \lambda^2\phi(t/\lambda) - \mu^2\frac{\omega(t/\lambda)}{\omega(t/\mu)}\phi(t/\mu). \qquad (13)$$

Note that $\alpha(t) \in [0,1]$ by the properties of ω. If we are able to invert $w = \alpha(t)$ in $[0,1]$, then $\gamma(w) = \beta(\alpha^{-1}(w))$ is given by

$$\gamma(w) = \lambda^2\phi(\alpha^{-1}(w)/\lambda) - \mu^2 w\,\phi\left(\frac{\alpha^{-1}(w)}{\mu}\right), \qquad (14)$$

and the lifted representation of $\phi(\cdot;\lambda)$ w.r.t. $\phi(\cdot;\mu)$ consequently reads as $\mu^2 w\,\phi(x/\mu) + \gamma(w)$. As it is shown in the following proposition, several robust kernels, among them the Welsch kernel $\phi_{\text{Wel},\tau}(\cdot)$, the smooth truncated kernel $\phi_{\text{ST},\tau}(\cdot)$ and the Geman-McClure kernel $\phi_{\text{Gem},\tau}(\cdot)$ (see Table 1), turn out to be very convenient for lifting w.r.t. a scaled version of themselves.

Proposition 3. Let $s \geq 1$, then the following identities hold:

1. $\phi_{\text{Wel},\tau}(x) = \min_{w \in [0,1]} w\phi_{\text{Wel},s\tau}(x) + \gamma_{\text{Wel},\tau,s}(w)$ with

$$\gamma_{\text{Wel},\tau,s}(w) = \frac{\tau^2}{2}\left(1 + w\left((s^2 - 1)w^{\frac{1}{s^2-1}} - s^2\right)\right),$$

2. $\phi_{\text{ST},\tau}(x) = \min_{w \in [0,1]} w\phi_{\text{ST},s\tau}(x) + \gamma_{\text{ST},\tau,s}(w)$ with $\gamma_{\text{ST},\tau,s}(w) = \frac{s^2\tau^2(w-1)^2}{4(s^2-w)}$,

3. $\phi_{\text{Gem},\tau}(x) = \min_{w \in [0,1]} w\phi_{\text{Gem},s\tau}(x) + \gamma_{\text{Gem},\tau,s}(w)$ with $\gamma_{\text{Gem},\tau,s}(w) = \frac{s^2\tau^2(\sqrt{w}-1)^2}{2(s^2-1)}$.

The proof is given in the supplementary material. Let us note that for each of these three kernels, $\lim_{s \to \infty} \gamma_{\cdot,\tau,s}(w)$ corresponds to the respective function $\gamma(w)$ in the half-quadratic lifting case (see Table 1). A graphical illustration of that proposition is given in Figure 1.

In the next section, we will leverage these propositions to show that any of these kernels can be optimized at a desired shape parameter τ by an iterated lifted representation using successively larger shape parameters.

4 Iterated Lifting

The half-quadratic lifting technique uses a quadratic kernel as basis kernel to represent a robust cost (see Eq. 1). Quadratic kernels are convenient because they enable efficient block-coordinate methods [2] and Gauss-Newton approaches [11]. However, as we showed in Figure 1, it may be beneficial to use a different (non-quadratic) basis kernel. Assume we have a robust kernel ϕ_2 that has a lifted representation w.r.t. another robust kernel ϕ_1. Unless ϕ_1 is a quadratic kernel, minimizing ϕ_2 by lifting using ϕ_1 as basis kernel does usually not allow to leverage an efficient optimization algorithm. Hence it is sensible to replace ϕ_1 by its "half-quadratic lifted" representation. This scheme can be iterated, for instance by considering ϕ_2 as basis kernel to lift a kernel ϕ_3 which will produce a gradual lifting of ϕ_3 and allow utilization of efficient non-linear least squares solvers. In the following we make this construction more precise and analyze its properties.

Definition 4. Let functions $(\gamma_k)_{k \in 1:K}$ with $\gamma_k : [0,1] \to \mathbb{R}_0^+$ be given. We introduce the mappings

$$\tilde{\phi}^{(K)}(\mathbf{w};x) = \left(\prod_{k=1}^{K} w_k\right)\frac{x^2}{2} + \sum_{k=1}^{K}\left(\prod_{l=k+1}^{K} w_l\right)\gamma_k(w_k) \tag{15}$$

$$\phi^{(K)}(x) = \min_{\mathbf{w} \in [0,1]^K} \tilde{\phi}^{(K)}(\mathbf{w};x). \tag{16}$$

We call $\tilde{\phi}^{(K)}$ the K-lifted representation of the induced robust kernel $\phi^{(K)}$. $\phi^{(K)}$ depends on γ_k for all k, but for brevity we omit this dependence, and the choice of $\{\gamma_k\}$ should be clear from the context.

Let $\phi_k : x \mapsto \min_{w \in [0,1]} wx^2/2 + \gamma_k(w)$ be a robust kernel. From Eq. 16, if we consider a 2-lifted representation, we obtain

$$\phi^{(2)}(x) = \min_{w_2 \in [0,1]} w_2 \min_{w_1 \in [0,1]}\left\{w_1\frac{x^2}{2} + \gamma_1(w_1)\right\} + \gamma_2(w_2) \tag{17}$$

$$= \min_{w_2 \in [0,1]} w_2\phi_1(x) + \gamma_2(w_2) = \phi_2\left(\sqrt{2\phi_1(x)}\right) = (\phi_2 \circ \zeta_1)(x) = \frac{1}{2}\left((\zeta_2 \circ \zeta_1)(x)\right)^2$$

where we used the contraction ζ associated with a robust kernel ϕ (see Lemma 1). We can generalize Eq. 17 to arbitrary K by induction over K and arrive at the following result for K-lifted representations,

Proposition 4. An explicit expression for $\phi^{(K)}$ is given by

$$\phi^{(K)}(x) = (\phi_K \circ \zeta_{K-1} \circ \ldots \circ \zeta_1)(x) = \frac{1}{2}\left((\zeta_K \circ \ldots \circ \zeta_1)(x)\right)^2. \tag{18}$$

Further, $\phi^{(K)}(x)$ is more robust than $\phi^{(K-1)}(x) = \phi_{K-1} \circ \zeta_{K-2} \circ \ldots \circ \zeta_1$ in the following sense: for all $x \in \mathbb{R}$ it holds that $\phi^{(K)}(x) \leq \phi^{(K-1)}(x)$ and $\omega^{(K)}(x) \leq \omega^{(K-1)}(x)$.

Proof. Eq. 18 can be seen by repeating the steps leading to Eq. 17, i.e. recursively applying $\phi_k(\cdot) = \min_{w'} w'x^2/2 + \gamma_k(w')$ for $k = 1, \ldots, K$ on Eq. 16, and using the definition of ζ_k. We also deduce that $\phi^{(K)}(x) \leq \phi^{(K-1)}(x)$ by applying Lemma 1. The final claim can be seen as follows: we recursively define $\bar{x}_0 = x$ and $\bar{x}_k = \zeta_k(\bar{x}_{k-1})$. Then by the chain rule the weight function $\omega^{(K)}$ can be written as

$$\omega^{(K)}(x) = \prod_{k=1}^{K} \frac{\phi_k'(\bar{x}_{k-1})}{\bar{x}_{k-1}} = \prod_{k=1}^{K} \omega_k(\bar{x}_k). \tag{19}$$

Hence we obtain $\omega^{(K)}(x) = \omega_K(\bar{x}_{K-1})\,\omega^{(K-1)}(x) \leq \omega^{(K-1)}(x)$ since $\omega_K(\bar{x}_{K-1}) \in [0,1]$. $\square$

The explicit shape of the weight function in Eq. 19 allows to analyze the limit function $\phi^{(K)}$ for $K \to \infty$ easily:

Proposition 5. Let a robust kernel ϕ be given, and let $\tau \stackrel{\text{def}}{=} \max\{x : \omega(x) = 1\}$ be its radius of unconditional inliers. Then

$$\phi^{(\infty)} = \lim_{K \to \infty} \phi \circ \underbrace{\zeta \circ \cdots \circ \zeta}_{K-1 \text{ times}} \quad \text{is given by} \quad \phi^{(\infty)}(x) = \frac{1}{2}\left(\min\{|x|, \tau\}\right)^2. \tag{20}$$

We can summarize the findings above as follows:

- A robust kernel can be interpreted as applying a non-linear contraction on the original residual, and iterated lifting amounts to chaining contractions.

- Consequently, iterating the lifting step to obtain a K-lifted representation leads to successively more robust kernels.

We finally illustrate our results of how to use iterated lifting for the Welsch kernel and $K = 3$ (using Prop. 3) as follows:

Proposition 6. We have the identity

$$\phi_{\text{Wel},\tau}(x) = \min_{w_1,w_2,w_3} w_3\left(w_2\left(w_1\frac{x^2}{2} + \gamma_{\text{Wel},s^2\tau,\infty}(w_1)\right) + \gamma_{\text{Wel},s\tau,s}(w_2)\right) + \gamma_{\text{Wel},\tau,s}(w_3) \tag{21}$$

Thus, we can optimize the Welsch kernel at a desired shape parameter τ by applying a non-linear least squares solver to an iterated lifted representation using successively larger shape parameters. Similar expressions can be obtained for other robust kernels such as the Smooth truncated kernel or Geman-McClure (again using Prop. 3).
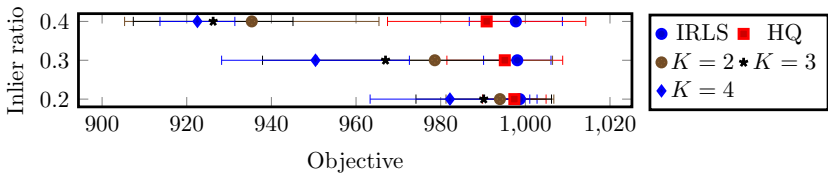
Figure 2: Average objective (and standard deviation) reached by the different methods for robustly fitting the mean to synthetically generated data points.

Implementation: The obvious choice for a numerical implementation is to minimize Eq 17 or Eq. 21 by jointly optimizing over all weights $w_1, \ldots, w_K$ and parameters θ. It can be shown theoretically and in practice, that this behaves very similar to standard half-quadratic lifting using joint optimization. The benefit of iterated lifting is the higher flexibility of choosing a schedule of how parameters and weights are updated. A schedule working well in our problem instances is given as follows: we increase the set of optimized variables after T Levenberg-Marquardt iterations (first θ only, then $w_1 \cup \theta$, then $w_2 \cup w_1 \cup \theta$ until $w_K \cup \ldots \cup w_1 \cup \theta$, and starting over the cycle). All weights $w_1 \ldots, w_K$ are initialized to 1. We use $T = 1$ in our experiments. Since weights have to be non-negative, we parametrize $w_k = u_k^2$ with $u_k \in \mathbb{R}$.

5 Numerical Results

We implemented Iterated Reweighted Least Squares (IRLS), joint half-quadratic optimization (HQ) and iterated lifting based on a sparse Levenberg-Marquardt method in C++ (using colamd and sparse Cholesky decomposition from the SuiteSparse library). Since we employ a sparse but direct solver for the trust region subproblems, our implementation is limited to medium large problem instances. We use the Welsch kernel as target cost and set $s = 2$ in the K-lifted representation.

Synthetic data: For the first experiment we robustly fit a mean vector θ to synthetically generated D-dimensional point data d at a given inlier ratio. Outliers are uniformly distributed in $[-20, 20]^D$, and inliers are in $\mathcal{N}(\mu, I_D)$, where the mean vector and its initial estimate θ are also uniformly sampled from $[-20, 20]^D$. The robust kernel is chosen to be the Welsch kernel at $\tau = 1/2$, thus the problem we consider has the following objective,

$$\min_{\theta} \sum_{i=1}^{N} \phi_{\text{Wel}, \frac{1}{2}} (\sqrt{(d_i - \theta)^T (d_i - \theta)}). \tag{22}$$

In order to apply iterated lifting to this problem, we simply use the expression given in Prop. 3 to obtain a K-lifted representation. For the case $K = 3$, it corresponds to applying Eq. 21 to transform Eq. 22 into a non-linear least squares problem. Fig. 2 depicts the average objective values (and the respective standard deviation) reached by the different methods after 100 runs for varying choices of inlier ratios and $D = 3$. There is a clear ordering of methods that can be summarized as 4-lifting $\succ$ 3-lifting $\succ$ 2-lifting $\succ$ HQ $\succ$ IRLS.
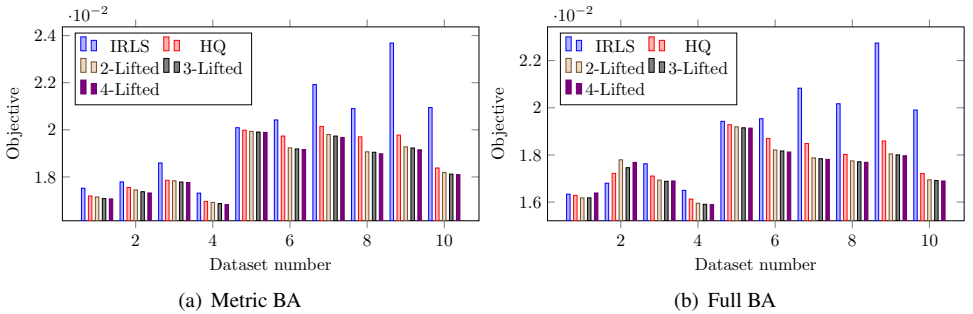
Figure 3: Objective values reached by the different methods for metric bundle adjustment (a) and full bundle adjustment additionally optimizing over focal length and lens distortion parameters (b).

Bundle adjustment data: We selected 2 problem instances from each of the 5 models in the publicly available bundle adjustment data set [1] and used the given camera parameters and 3D structure as initializer. The bundle adjustment objective is given by

$$\sum \phi_{\mathrm{Wel}, \frac{1}{2}} \big(f_i \eta_i \left(\pi(R_i X_j + t_i) \right) - \hat{p}_{ij} \big), \tag{23}$$

where $\hat{p}_{ij} \in \mathbb{R}^2$ is the observed image observation of the j-th 3D $X_j \in \mathbb{R}^3$ point in the i-image with orientation $R_i \in SO(3)$ and $t_i \in \mathbb{R}^3$ and focal length f_i. $\pi : \mathbb{R}^3 \to \mathbb{R}^2$, $\pi(X) = X/X_3$ is the projection function, and η_i is the lens distortion function with $\eta_i(p) = (1 + k_{i,1}\|p\|^2 + k_{i,2}\|p\|^4)p$. Fig 3(a) and (b) illustrate the achieved objective values using different methods for robust cost optimization for metric (optimize over 3D structure $\{X_j\}$ and camera poses $\{(R_i, t_i)\}$) and full (optimize over f_i and lens distortion parameters as well) bundle adjustment. In our current straightforward implementation 3-lifting has approximately 25% overhead compared to the scheme proposed in [11], but reaches significantly better minima in almost all cases (except dataset 2 "Ladybug-598", which leads to "flatlining" of the objective—i.e. large values for the damping parameter and extremely small improvements—for several methods). We refer to the supplementary material for additional results.

6   Conclusion

In this paper, we proposed a novel robust cost optimization strategy called *iterated lifting* that provides a gradual lifting scheme for a target robust kernel. To do so, we generalized the standard half-quadratic lifting construction to non-quadratic basis kernels in order to be able to lift a target robust kernel using another, less robust kernel as basis kernel. Then, we demonstrated that this process can be iterated until the quadratic kernel is reached to allow utilization of efficient non-linear least-squares solvers. The performances of our novel approach were evaluated against Iterated Reweighted Least Squares (IRLS) [6] and joint half-quadratic optimization (HQ) [11] both on synthetic and real data. In both cases, our approach reached lower objective values than IRLS and HQ. As a future work, we plan to investigate different lifting strategies that also allow a gradual lifting for a target robust kernel.

References

[1] Sameer Agarwal, Noah Snavely, Steven M Seitz, and Richard Szeliski. Bundle adjustment in the large. In *Proc. ECCV*, pages 29–42. Springer, 2010.

[2] Michael J Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1):57–91, 1996.

[3] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proc. CVPR*, pages 5556–5565, 2015.

[4] Donald Geman and George Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383, 1992.

[5] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, (6):721–741, 1984.

[6] Peter J Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192, 1984.

[7] Philipp Heise, Brian Jensen, Sebastian Klose, and Alois Knoll. Variational patchmatch multiview reconstruction and refinement. In *Proc. ICCV*, pages 882–890, 2015.

[8] Jérôme Idier. Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *IEEE Trans. Image Proc.*, 10(7):1001–1009, 2001.

[9] Andrew Jaegle, Stephen Phillips, and Kostas Daniilidis. Fast, robust, continuous monocular egomotion computation. In *Proc. ICRA*, pages 773–780. IEEE, 2016.

[10] Niko Sünderhauf and Peter Protzel. Switchable constraints for robust pose graph slam. In *Proc. IROS*, pages 1879–1884. IEEE, 2012.

[11] Christopher Zach. Robust bundle adjustment revisited. In *Proc. ECCV*, pages 772–787, 2014.

[12] Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, et al. Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transactions on Graphics (TOG)*, 33(4):156, 2014.