Descending, lifting or smoothing: Secrets of robust cost optimization

Christopher Zach (christopher.m.zach@gmail.com) — Toshiba Research Europe, Cambridge, UK

Guillaume Bourmaud (guillaume.bourmaud@u-bordeaux.fr) — University of Bordeaux, France

TOSHIBA Leading Innovation >>> — université de BORDEAUX

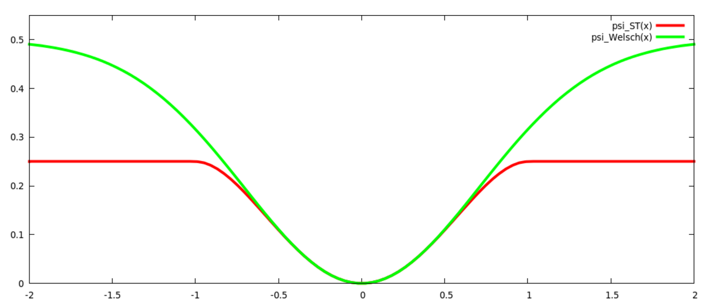
1) Introduction

Problem statement

Minimize a cost function involving robust data terms

$$\min_{\boldsymbol\theta} \Psi(\boldsymbol\theta) \quad \text{with} \quad \Psi(\boldsymbol\theta) = \sum_i \psi(\|\mathbf{f}_i(\boldsymbol\theta)\|)$$

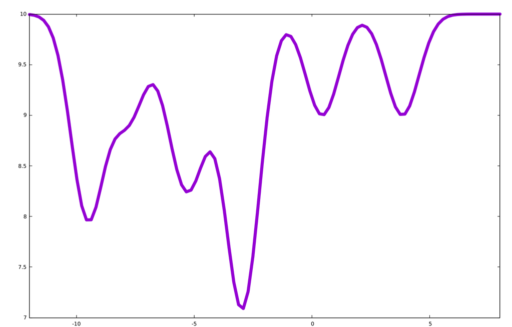
where $\mathbf{f}_i(\boldsymbol\theta) : \mathbb{R}^p \to \mathbb{R}^d$ and $\psi(\cdot)$ is a robust kernel function.

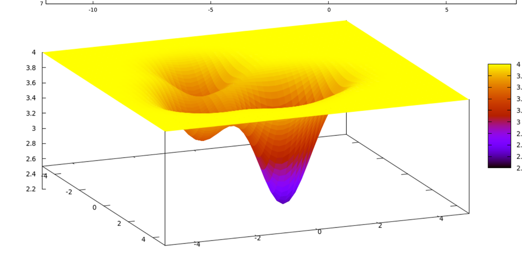


Challenges

- large number of local minima
- large number of parameters to estimate

How to obtain an efficient algorithm able to escape poor local minima?





Take-home message:

	Direct methods	Lifting methods	Graduated optimization
Speed of convergence	++	+	-
Ability to escape local minima	-	+	++

2) Direct methods

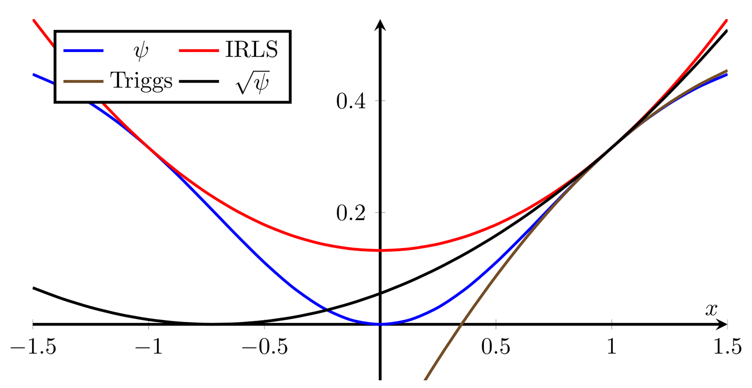
Minimize $\Psi(\boldsymbol\theta)$ directly using an NLLS solver:

1. First order approximation of $\mathbf{f}_i(\boldsymbol\theta)$ at $\boldsymbol\theta = \bar{\boldsymbol\theta}$:

$$\mathbf{f}_i(\bar{\boldsymbol\theta} + \Delta\boldsymbol\theta) \approx \bar{\mathbf{f}}_i + \mathtt{J}_i \Delta\boldsymbol\theta$$

2. Approximate $\psi(\|\bar{\mathbf{f}}_i + \mathtt{J}_i\Delta\boldsymbol\theta)\|)$ with a quadratic model $\check\psi$ s.t.

$$\psi(\|\bar{\mathbf{f}}_i\|) = \check\psi(\|\bar{\mathbf{f}}_i\|)$$

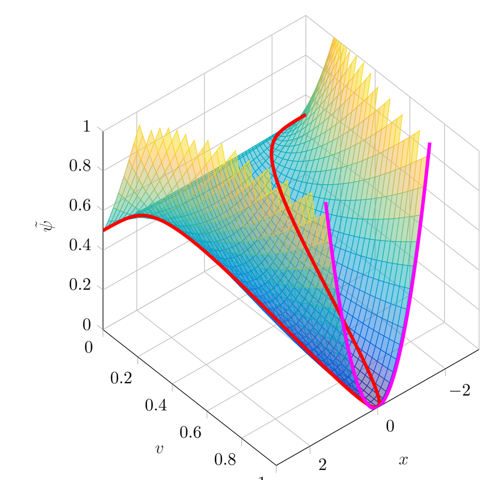


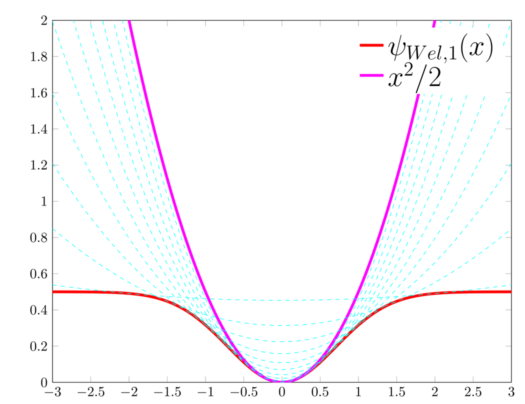
Quadratic surrogate models used by different direct approaches at $\bar\theta = 1$ (the x-axis corresponds to $\theta = \bar\theta + \Delta\theta$)

Our contribution The underlying quadratic models are very different and the IRLS model has desirable properties.

3) Half-quadratic lifting-based methods

1. Use a *quadratic basis kernel* to lift $\psi(x)$: $$\psi(x) = \min_{v\in[0,1]} \frac{v}{2}x^2 + \gamma(v)$$





Half-quadratic lifting [1] of $\psi_{\mathrm{Wel},1}(x)$,

2. Rewrite $\Psi(\boldsymbol\theta)$ to obtain an NLLS problem:

$$\min_{\boldsymbol\theta} \sum_i \psi(\|\mathbf{f}_i(\boldsymbol\theta)\|) = \min_{\boldsymbol\theta,\{v_i\}_i} \sum_i \left\| \begin{matrix} \frac{\sqrt{v_i}}{\sqrt{2}}\mathbf{f}_i(\boldsymbol\theta) \\ \sqrt{\gamma(v_i)} \end{matrix} \right\|^2$$

Our contribution A convexified Newton approximation to replace the Gauss-Newton approximation [1] in the NLLS solver.

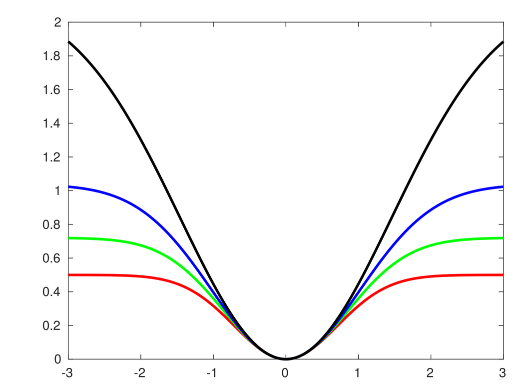
[1] Christopher Zach. Robust bundle adjustment revisited. ECCV 2014.

4) Graduated optimization

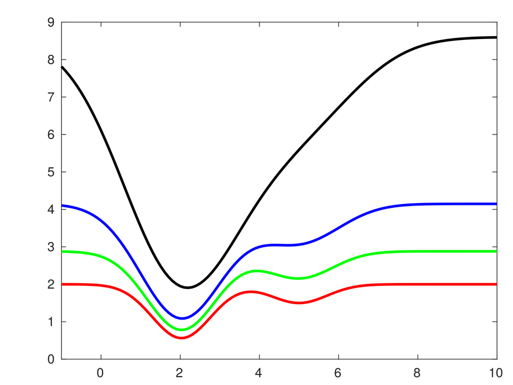
1. Build a sequence of objectives $(\Psi^0, \dots, \Psi^{k_{\max}})$ s.t. $\Psi^0 = \Psi$ and Ψ^{k+1} is "easier" to optimize than Ψ^k

   Natural approach: $(s_k)_{k=0}^{k_{\max}}$ s.t. $s_0 = 1$ and $s_k < s_{k+1}$,

$$\Psi^k(\boldsymbol\theta) := \sum_i \psi^k(\|\mathbf{f}_i(\boldsymbol\theta)\|) \text{ with } \psi^k(r) := s_k^2\psi(r/s_k)$$



Sequence of smoothed kernels (ψ^k)



Sequence of objectives (Ψ^k)

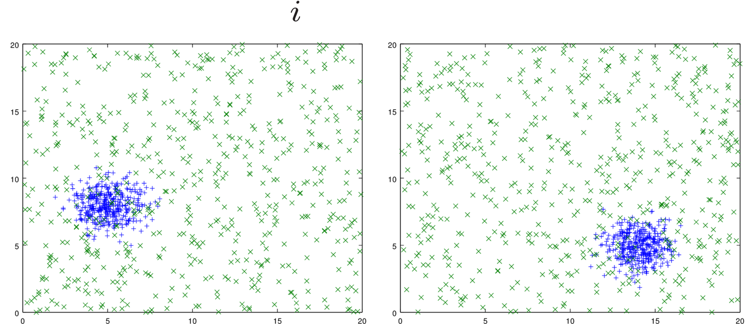
2. Successively optimize the sequence of cost functions using for instance a direct solver

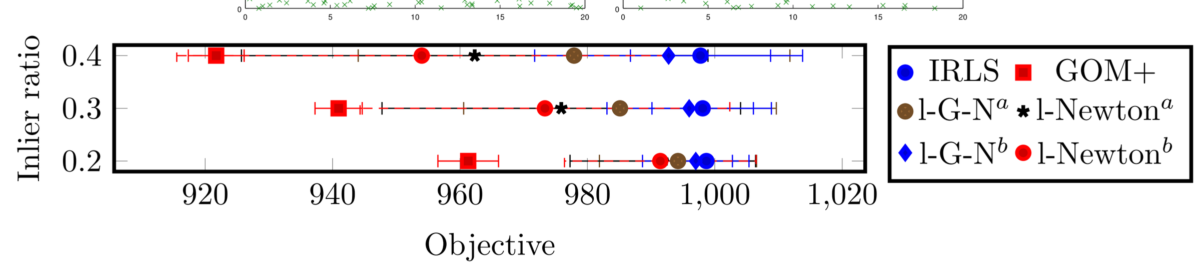
Our contribution A novel stopping criterion to speed-up graduated optimization methods.

5) Results

Robust mean

$$\min_{\boldsymbol\theta} \sum_i \psi(\|\boldsymbol\theta - \mathbf{y}_i\|)$$





Average objective (and standard deviation) reached by the different methods for the "robust mean" problem

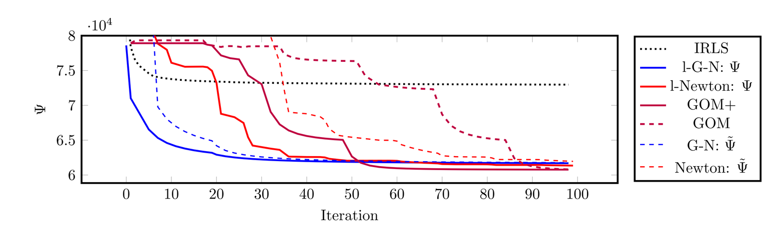
Image smoothing

$$\min_{\boldsymbol\theta} \sum\nolimits_{i\in\mathcal{V}} \psi^{\text{data}}(\theta_i - u_i) + \sum\nolimits_{(i,j)\in\mathcal{E}} \psi^{\text{smooth}}(\theta_i - \theta_j)$$

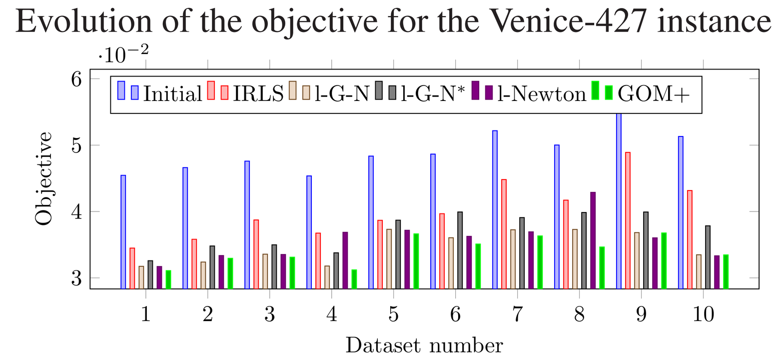
Visual results for the "image smoothing" problem using random init

Bundle adjustment

$$\min_{\{\mathtt{R}_i,\mathbf{t}_i\}_i,\{\mathbf{X}_j\}_j} \sum\nolimits_{i,j} \psi\left(\|\pi(\mathtt{R}_i\mathbf{X}_j + \mathbf{t}_i) - \mathbf{q}_{ij}\|\right)$$



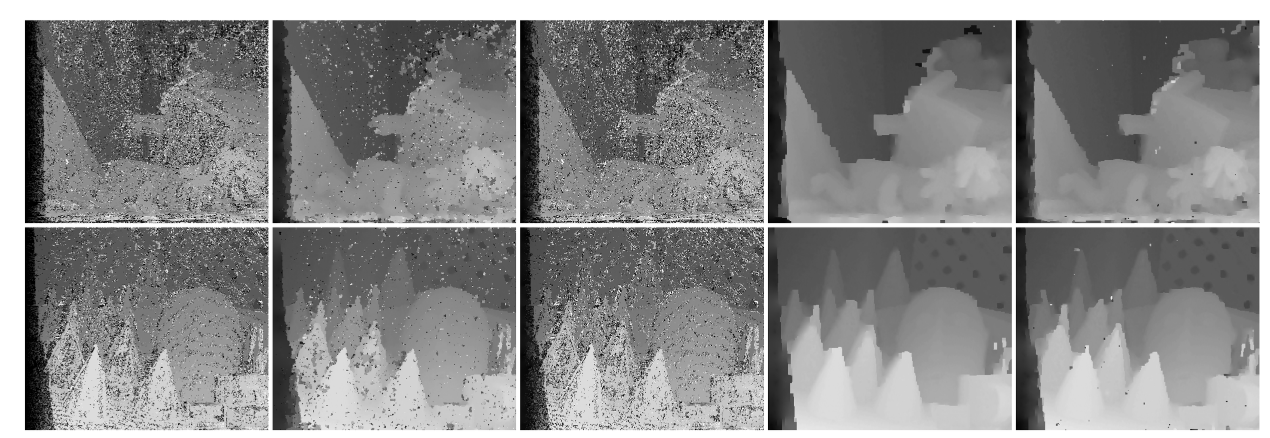
Evolution of the objective for the Venice-427 instance



Objective values reached by the different methods for metric BA

Dense stereo

$$\min_{\boldsymbol\theta} \frac{\lambda}{2} \sum\nolimits_{i\in\mathcal{V}} \sum_{k=1}^{K} \psi^{\text{data}}(\theta_i - d_{i,k}) + \sum\nolimits_{(i,j)\in\mathcal{E}} \psi^{\text{smooth}}(\theta_i - \theta_j)$$



Visual results for the "variational stereo" problem using random init