Global Convergence of Alternating Direction Method of Multipliers for Invex Objective Losses

Science and Technology Facilities Council

Samuel Pinilla, Siu-lun Yeung, Jeyan T.

Introduction

The class of functions that any stationary point is a global minimizer is defined as follows.

Definition (Invexity). *Let $f:\mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz; then f is invex if there exists a function $\eta:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ such that*

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \zeta^T \eta(\boldsymbol{x},\boldsymbol{y}),$$

$\forall \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n, \forall\zeta\in\partial f(\boldsymbol{y})$.

Hierarchy of optimizable functions

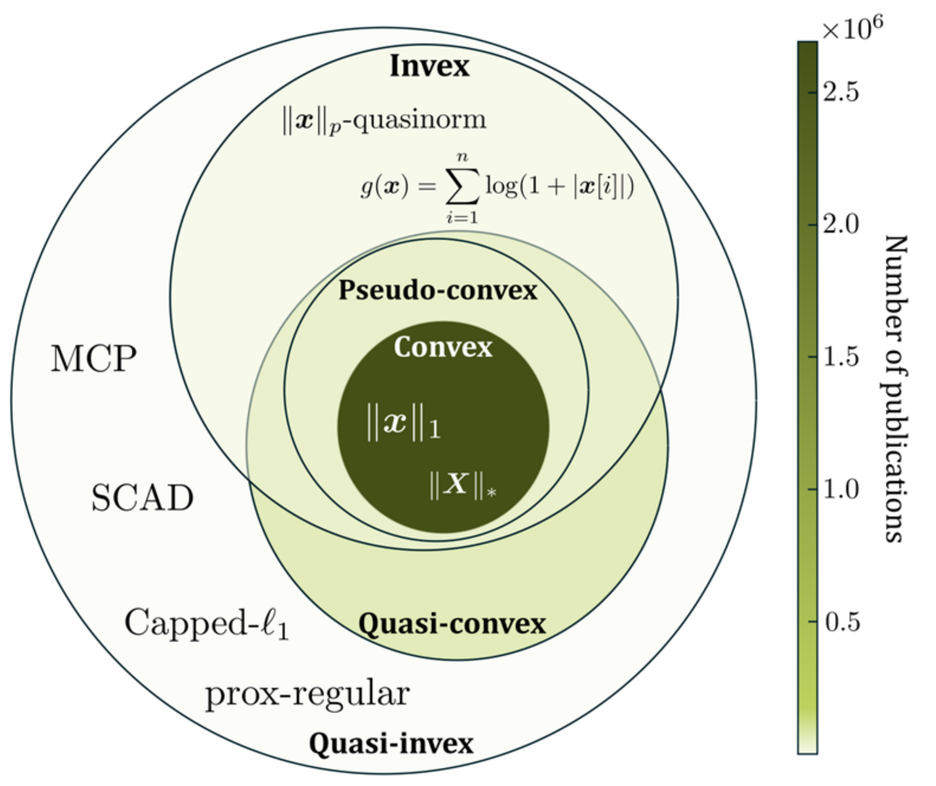

Fig 1. Our contribution is identifying invex and quasi-invex functions relevant for imaging applications.

Background

A reconstruction task is the solution of:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{z}\in\mathbb{R}^p}{\text{minimize}}\ f(\boldsymbol{x})+g(\boldsymbol{z})$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}=\boldsymbol{y}$$

where $\boldsymbol{A}\in\mathbb{R}^{m\times n}, \boldsymbol{B}\in\mathbb{R}^{m\times p}$, and $\boldsymbol{y}\in\mathbb{R}^m$. In order to solve it, the Alternating Direction Method of Multipliers is used.

Limitations

- Global guarantees of ADMM are not available for non-convex mappings.
- Global guarantees of ADMM were extended to prox-regular functions.
- Prox-regular functions do not ensure global minima

Material and Methods

Proposed family of functions

Definition Let $h:\mathbb{R}^n\to\mathbb{R}$ such that $h(\boldsymbol{x})=\sum_{i=1}^n s(|\boldsymbol{x}[i]|)$, where $s:[0,\infty)\to[0,\infty)$ and $s'(w)>0$ for $w\in(0,\infty)$. If s with $s(0)=0$ such that $s(w)/w^2$ is non-increasing on $(0,\infty)$, then $h(\boldsymbol{x})$ is said to be an *admissible function*.

Properties of proposed family of functions

Theorem 1. Let $f,g:\mathbb{R}^n\to\mathbb{R}$ be two admissible functions as in Definition , such that $f(\boldsymbol{x})=\sum_{i=1}^n s_f(|\boldsymbol{x}[i]|)$, and $g(\boldsymbol{x})=\sum_{i=1}^n s_g(|\boldsymbol{x}[i]|)$. Then the following holds:

- $f(\boldsymbol{x})$, and $g(\boldsymbol{x})$ are invex;
- $h(\boldsymbol{x})=\alpha f(\boldsymbol{x})+\beta g(\boldsymbol{x})$ is an admissible function (therefore invex) for every $\alpha,\beta\geq 0$;
- $h(\boldsymbol{x})=\sum_{i=1}^n (s_f\circ s_g)(|\boldsymbol{x}[i]|)$ is admissible function.
- $h(\boldsymbol{x})=\sum_{i=1}^n \min(s_f(|\boldsymbol{x}[i]|), s_g|\boldsymbol{x}[i]|)$ is admissible function.
- $h(\boldsymbol{x})=\sum_{i=1}^n \max(s_f(|\boldsymbol{x}[i]|), s_g|\boldsymbol{x}[i]|)$ is admissible function.

Examples of invex functions

Theorem 2. All the following functions for $c,\delta>0$, and $\alpha\in\mathbb{R}$ are admissible

$$h(\boldsymbol{x})=\sum_{i=1}^n\log\left(1+\frac{\boldsymbol{x}^2[i]}{\delta^2}\right) \quad (5)$$

$$h(\boldsymbol{x})=\sum_{i=1}^n\frac{2\boldsymbol{x}^2[i]}{\boldsymbol{x}^2[i]+4\delta^2} \quad (6)$$

$$h(\boldsymbol{x})=\sum_{i=1}^n\frac{|\alpha-2|}{\alpha}\left(\left(\frac{(\boldsymbol{x}[i]/c)^2}{|\alpha-2|}+1\right)^{\alpha/2}-1\right) \quad (7)$$

$$h(\boldsymbol{x})=\sum_{i=1}^n\log\left(1+\boldsymbol{x}^2[i]\right)-\frac{\boldsymbol{x}^2[i]}{2\boldsymbol{x}^2[i]+2} \quad (8)$$

ADMM algorithm

$$\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v})=f(\boldsymbol{x})+g(\boldsymbol{z})+\frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}-\boldsymbol{y}+\boldsymbol{v}\|_2^2,$$

where $\boldsymbol{v}\in\mathbb{R}^m$ is the dual variable, and $\rho>0$. The optimization of $\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v})$ is summarized as

$$\boldsymbol{x}^{(t+1)} := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\left(f(\boldsymbol{x})+\frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}^{(t)}-\boldsymbol{y}+\boldsymbol{v}^{(t)}\|_2^2\right)$$

$$\boldsymbol{z}^{(t+1)} := \underset{\boldsymbol{z}\in\mathbb{R}^p}{\arg\min}\left(g(\boldsymbol{z})+\frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}^{(t+1)}+\boldsymbol{B}\boldsymbol{z}-\boldsymbol{y}+\boldsymbol{v}^{(t)}\|_2^2\right)$$

$$\boldsymbol{v}^{(t+1)} := \boldsymbol{v}^{(t)}+\boldsymbol{A}\boldsymbol{x}^{(t+1)}+\boldsymbol{B}\boldsymbol{z}^{(t+1)}-\boldsymbol{y}.$$

Results and Experiments

We evaluate the utility of the proposed family of invex functions to solve a Total Variation regularization problem.

Convergence guarantees

Theorem 3. Let $f(\boldsymbol{x}),g(\boldsymbol{z})$ be any invex construct in Theorem 1, with $\rho\sigma_n(\boldsymbol{A})\geq 1$, and $\rho\sigma_p(\boldsymbol{B})\geq 1$ (maximum singular values of $\boldsymbol{A}$, and $\boldsymbol{B}$ respectively). Assume $\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v})$ has a saddle point, that is, there exists $(\boldsymbol{x}^*,\boldsymbol{z}^*,\boldsymbol{v}^*)$ for which

$$\mathcal{L}_\rho(\boldsymbol{x}^*,\boldsymbol{z}^*,\boldsymbol{v})\leq\mathcal{L}_\rho(\boldsymbol{x}^*,\boldsymbol{z}^*,\boldsymbol{v}^*)\leq\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v}^*),$$

for all $\boldsymbol{x},\boldsymbol{z}$, and $\boldsymbol{v}$. Then

- Residual $\|\boldsymbol{r}^{(t)}\|_2=\|\boldsymbol{A}\boldsymbol{x}^{(t)}+\boldsymbol{B}\boldsymbol{z}^{(t)}-\boldsymbol{y}\|_2\to 0$;
- $\boldsymbol{v}^{(t)}\to\boldsymbol{v}^*$ as $t\to\infty$ where $\boldsymbol{v}^*$ is the dual optimal point;
- $f(\boldsymbol{x}^{(t)})+g(\boldsymbol{z}^{(t)})\to f(\boldsymbol{x}^*)+g(\boldsymbol{z}^*)$.

Additionally, the convergence rate is $\mathcal{O}(1/t)$

Numerical experiments

Table 1: Performance Results: Best: green, Second best: yellow, and the worst: red.

$f(\boldsymbol{x})$	Metrics	$g(\boldsymbol{z})$: ℓ_p	Log	Log-sub	SCAD	ℓ_1-norm
Eq. (5)	SSIM	0.6403	0.6267	0.6231	0.6195	0.6159
	MS-SSIM	0.9344	0.9296	0.9249	0.9202	0.9156
	ADMM-residual	$8.8\cdot10^{-4}$	$1.1\cdot10^{-3}$	$1\cdot10^{-3}$	$1.3\cdot10^{-3}$	$1.5\cdot10^{-3}$
Eq. (6)	SSIM	0.6361	0.6230	0.6166	0.6295	0.6104
	MS-SSIM	0.9289	0.9208	0.9168	0.9248	0.9128
	ADMM-residual	$9.4\cdot10^{-4}$	$1.3\cdot10^{-3}$	$1.5\cdot10^{-3}$	$1.1\cdot10^{-3}$	$1.9\cdot10^{-3}$
Eq. (7)	SSIM	0.6488	0.6378	0.6432	0.6324	0.6271
	MS-SSIM	0.9455	0.9331	0.9393	0.9271	0.9211
	ADMM-residual	$8\cdot10^{-4}$	$8.8\cdot10^{-4}$	$8.4\cdot10^{-4}$	$9.4\cdot10^{-4}$	$1\cdot10^{-3}$
Eq. (8)	SSIM	0.6445	0.6327	0.6386	0.6270	0.6214
	MS-SSIM	0.9399	0.9290	0.9344	0.9236	0.9183
	ADMM-residual	$8.4\cdot10^{-4}$	$1\cdot10^{-3}$	$9\cdot10^{-4}$	$1.1\cdot10^{-3}$	$1.2\cdot10^{-3}$
ℓ_2-norm	SSIM	0.6320	0.6182	0.6250	0.6050	0.6115
	MS-SSIM	0.9235	0.9168	0.9201	0.9101	0.9134
	ADMM-residual	$1\cdot10^{-3}$	$1.5\cdot10^{-3}$	$1.2\cdot10^{-3}$	$2.8\cdot10^{-3}$	$1.9\cdot10^{-3}$

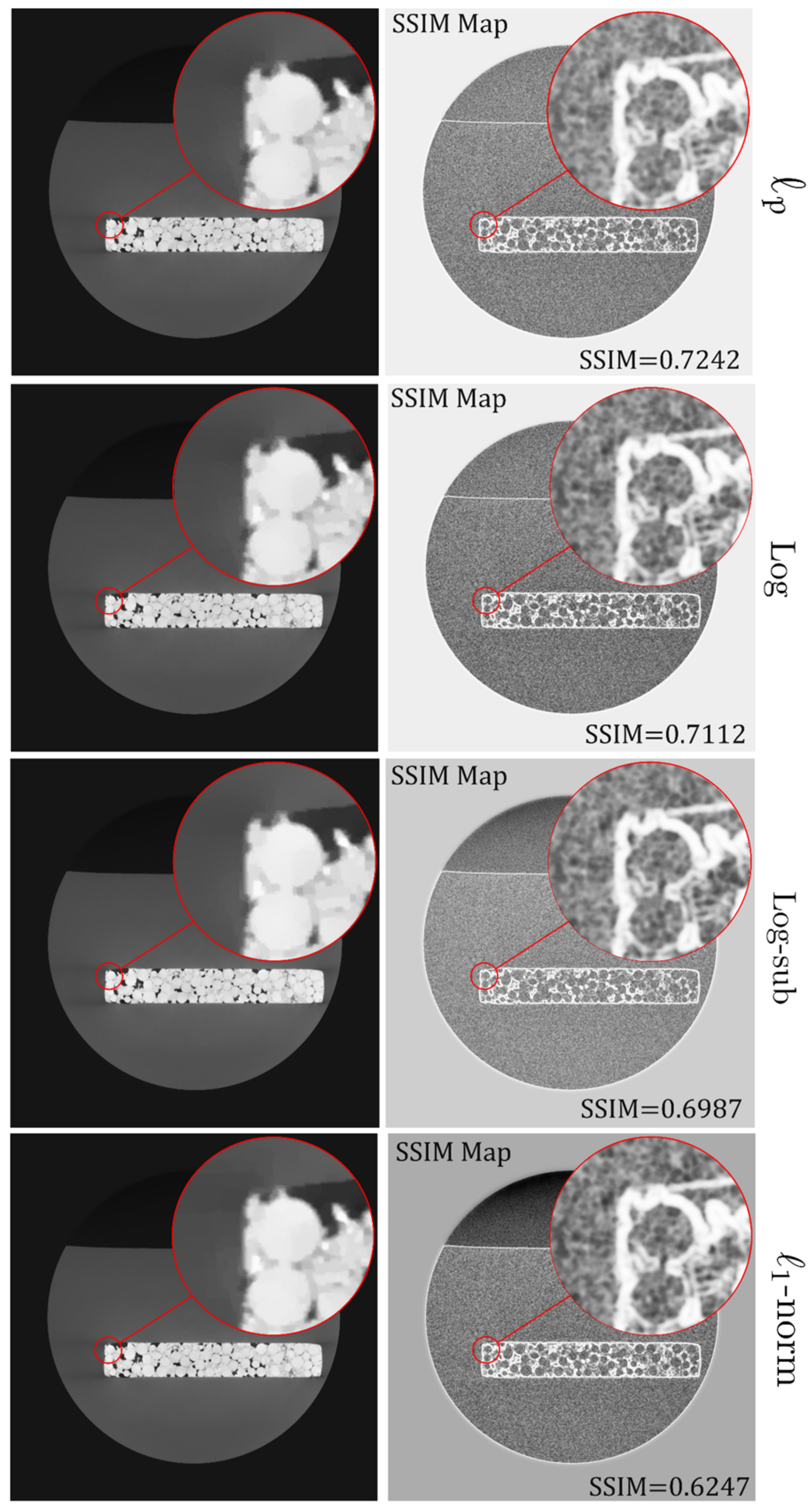

Fig 2. Restored images using the ADMM. We employ the SSIM map for each image and its averaged value.

Conclusion

- This paper identifies a family of functions for signal restoration.
- We provided the proof for the invex behaviours of these functions and global optimality with their convergence rate.
- This theoretical analysis to handle ADMM optimization problem, is first in its kind, and the approach is applicable to various other constrained optimization problems.

Acknowledgment

This work is partially supported by the EPSRC grant, Blueprinting for AI for Science at Exascale (BASE-II, EP/X019918/1), and by STFC Facilities Fund.

Contact Information

Email: samuel.pinilla@stfc.ac.uk

References

- Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," Journal of Scientific Computing.
- Pinilla, S., Thiyagalingam, J. Global Optimality for Non-linear Constrained Restoration Problems via Invexity. In The Twelfth International Conference on Learning Representations.