Iteratively regularized Newton methods with general data misfit functionals and applications to Poisson data

Frank Werner

Institute for numerical and applied mathematics
University of Göttingen, Germany

joint work with Thorsten Hohage

AIP 2011

Outline

Outline

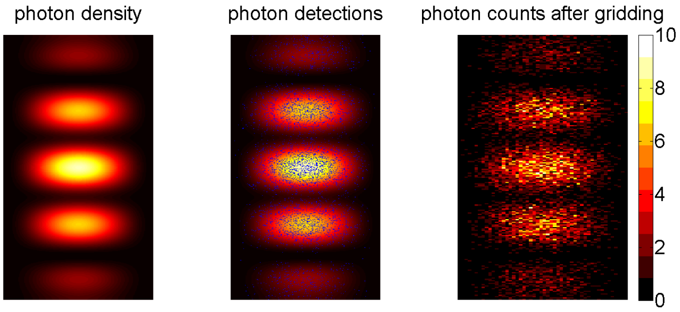
Photonic imaging

- Photonic imaging consists in counting photons which have interacted with some unknown object of interest.
- We want to reconstruct information on the unknown object $\varphi^\dagger$ contained in these photon counts.
- Formulation as an operator equation

$$F(\varphi) = g$$

  where g describes the photon density on the manifold where measurements are taken.
- For fundamental physical reasons, photon count data g^{obs} are Poisson distributed with mean $g^\dagger$ (true photon density).

Examples



- Positron Emission Tomography (PET)
- astronomical imaging
- scanning fluorescence microscopy, e.g. standard confocal, 4Pi or STED microscopy
- coherent x-ray imaging

Ill-Posedness

The forementioned problems are ill-posed in the sense that φ does not depend continuously on $F(\varphi)$. Hence, the problem cannot be solved directly or by a usual Newton method but regularization is needed.

For nonlinear F one of the most popular methods is the iteratively regularized Gauss-Newton method (IRGNM)

$$\varphi_{j+1} = \underset{\varphi \in \mathfrak{B}}{\operatorname{argmin}} \left(\left\| F'(\varphi_j; \varphi - \varphi_j) + F(\varphi_j) - g^{\mathrm{obs}} \right\|_{\mathbf{L}^2}^2 + \alpha_j \left\| \varphi - \varphi_0 \right\|_{\mathbf{L}^2}^2 \right)$$

with some initial guess $\varphi_0 \in \mathfrak{B}$.

The regularization parameters α_j control the stability and fulfill

$$\alpha_0 \leq 1, \qquad \alpha_j \searrow 0, \qquad 1 \leq \frac{\alpha_j}{\alpha_{j+1}} \leq C_{\mathrm{dec}} \qquad \text{for all} \qquad j \in \mathbb{N}.$$

Noise adjusted regularization

The IRGNM corresponds to a Gaussian noise structure. Hence,

- the information about the noise structure is ignored and
- especially for low intensity we get bad reconstructions.

Our idea is to use another data misfit functional $\mathcal{S}$ which incorporates the special structure of the noise and take

$$\varphi_{n+1} = \operatorname*{argmin}_{\varphi \in \mathfrak{B}} \mathcal{S}\left(F(\varphi_j) + F'(\varphi_j; \varphi - \varphi_j); g^{\mathrm{obs}}\right) + \alpha_j \mathcal{R}(\varphi)$$

where $\mathcal{S}\left(\cdot; g^{\mathrm{obs}}\right)$ is some convex data misfit functional and $\mathcal{R}$ some convex penalty term. For Poisson data the first choice would be the negative log-likelihood

$$\mathcal{S}\left(g; g^{\mathrm{obs}}\right) = \int_{\Omega} g - g^{\mathrm{obs}} \ln(g) \, \mathrm{d}x.$$

Alternatives

An alternative approach is nonlinear Tikhonov regularization

$$\varphi_\alpha = \underset{\varphi \in \mathfrak{B}}{\operatorname{argmin}} \, \mathcal{S}\left(F(\varphi); g^{\mathrm{obs}}\right) + \alpha \mathcal{R}(\varphi)$$

which has been considered by several authors:

J. M. Bardsley.
A theoretical framework for the regularization of Poisson likelihood estimation problems.
Inverse Problems and Imaging, 4:11–17, 2010.

M. Benning and M. Burger.
Error estimates for general fidelities.
Electronic Transactions on Numerical Analysis, 38:44–68, 2011.

J. Flemming.
Theory and examples of variational regularisation with non-metric fitting functionals.
Journal of Inverse and Ill-Posed Problems, 18(6):677–699, 2010.

O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen.
Variational Methods in Imaging.
Applied Mathematical Sciences. Springer, 2008.

Outline

Source condition I

The usual Hilbert space source condition

$$\varphi^\dagger - \varphi_0 = \Lambda\left(F'\left[\varphi^\dagger\right]^* F'\left[\varphi^\dagger\right]\right)\omega$$

implies by spectral theory and Jensen's inequality

$$\left|\left\langle \varphi_*^\dagger, \varphi - \varphi^\dagger \right\rangle\right| \leq \|\omega\| \left\|\varphi - \varphi^\dagger\right\| \Lambda\left(\frac{\left\|F'\left[\varphi^\dagger\right]\left(\varphi - \varphi^\dagger\right)\right\|^2}{\left\|\varphi - \varphi^\dagger\right\|^2}\right).$$

This is the prototype of a variational source condition.

B. Kaltenbacher and B. Hofmann.
Convergence Rates for the Iteratively Regularized Gauss-Newton Method in Banach Spaces.
Inverse Problems, 26(3):035007, 2010.

Source condition II

We will assume the following generalization:

Multiplicative variational source condition

There exists $\varphi_*^\dagger \in \partial\mathcal{R}\left(\varphi^\dagger\right) \subset \mathcal{X}'$, $\beta \geq 0$ and a concave index function $\Lambda : (0,\infty) \to (0,\infty)$ (i.e. continuous, monotonically increasing and $\Lambda(0) = 0$) such that

$$\left\langle \varphi_*^\dagger, \varphi^\dagger - \varphi \right\rangle \leq \beta \Delta\left(\varphi, \varphi^\dagger\right)^{\frac{1}{2}} \Lambda\left(\frac{\mathcal{S}\left(F\left(\varphi\right); g^\dagger\right)}{\Delta\left(\varphi, \varphi^\dagger\right)}\right) \qquad \text{for all } \varphi \in \mathfrak{B}.$$

Moreover, we assume that $t \mapsto \frac{\Lambda(t)}{\sqrt{t}}$ is monotonically decreasing.

$\Delta\left(\varphi, \varphi^\dagger\right) := \mathcal{R}\left(\varphi\right) - \mathcal{R}\left(\varphi^\dagger\right) - \left\langle \varphi_*^\dagger, \varphi - \varphi^\dagger \right\rangle$ is the Bregman distance.

Noise I

In case of $\mathcal{S}$ being the r-th power of a norm one usually assumes $\left\|g^{\mathrm{obs}} - g^\dagger\right\| \leq \delta$ which by the triangle inequality leads to

$$2^{1-r}\left\|g - g^\dagger\right\|^r - \delta^r \leq \left\|g - g^{\mathrm{obs}}\right\|^r \leq 2^{r-1}\left\|g - g^\dagger\right\|^r + 2^{r-1}\delta^r$$

for all $g \in \mathcal{Y}$.

In case of Poisson noise and the negative log-likelihood as data misfit, we obtain the following difficulties:

- The data misfit functional does not fulfill a triangle inequality.
- $\mathcal{S}\left(g; g^{\mathrm{obs}}\right)$ might be ∞ even if $\mathcal{S}\left(g; g^\dagger\right)$ is finite and vice versa.

Noise II

In case of $\mathcal{S}$ being the r-th power of a norm one usually assumes $\left\|g^{\mathrm{obs}} - g^\dagger\right\| \leq \delta$ which by the triangle inequality leads to

$$2^{1-r}\left\|g - g^\dagger\right\|^r - \delta^r \leq \left\|g - g^{\mathrm{obs}}\right\|^r \leq 2^{r-1}\left\|g - g^\dagger\right\|^r + 2^{r-1}\delta^r$$

for all $g \in \mathcal{Y}$.

Generalization:

Noise level

There exists some $C_{\mathrm{err}} \geq 1$ and a functional $\mathbf{err} : \mathcal{Y} \to [0, \infty]$ such that

$$\frac{1}{C_{\mathrm{err}}}\mathcal{S}\left(g; g^\dagger\right) - \mathbf{err}\left(g\right) \leq \mathcal{S}\left(g; g^{\mathrm{obs}}\right) \leq C_{\mathrm{err}}\mathcal{S}\left(g; g^\dagger\right) + C_{\mathrm{err}}\mathbf{err}\left(g\right)$$

for all $g \in \mathcal{Y}$.

Nonlinearity estimate

Generalized tangential cone condition

There exist constants η (later assumed to be sufficiently small) and $C_{\mathrm{tc}} \geq 1$ such that

$$\begin{aligned}
&\frac{1}{C_{\mathrm{tc}}}\mathcal{S}\left(F(\psi);g^\dagger\right) - \eta\mathcal{S}\left(F(\varphi);g^\dagger\right) \\
\leq &\mathcal{S}\left(F(\varphi) + F'(\varphi;\psi-\varphi);g^\dagger\right) \\
\leq &C_{\mathrm{tc}}\mathcal{S}\left(F(\psi);g^\dagger\right) + \eta\mathcal{S}\left(F(\varphi);g^\dagger\right) \qquad \text{for all } \varphi,\psi \in \mathfrak{B}.
\end{aligned}$$

For $\mathcal{S}(g;\hat{g}) = \|g-\hat{g}\|^r$ this follows from the standard *tangential cone condition*

$$\left\|F(\varphi) - F(\psi) - F'(\varphi;\psi-\varphi)\right\| \leq \bar{\eta}\left\|F(\varphi) - F(\psi)\right\|.$$

Rate function and stopping rule

Our convergence rates result uses the following rate function:

$$\Theta(t) := t\Lambda^2(t).$$

Θ and Θ^{-1} are index functions.

Rate function and stopping rule

Our convergence rates result uses the following rate function:

$$\Theta(t) := t\Lambda^2(t).$$

Θ and Θ^{-1} are index functions. Moreover define

$$\begin{aligned} \mathbf{err}_j &:= \mathbf{err}\left(F(\varphi_j) + F'(\varphi_j; \varphi_{j+1} - \varphi_j)\right) \\ &\quad + C_{\mathrm{err}}\,\mathbf{err}\left(F(\varphi_j) + F'\left(\varphi_j; \varphi^\dagger - \varphi_j\right)\right) \end{aligned}$$

and use the following stopping index:

Stopping rule

We define

$$j_*(\mathbf{err}_j) := \min\left\{ j \in \mathbb{N} \;\middle|\; \Theta(\alpha_j) \leq \tau\,\mathbf{err}_j \right\}$$

with some tuning parameter $\tau \geq 1$.

Rates of convergence

Convergence theorem

Let the Assumptions from above hold and let η, $\Delta\left(\varphi_0, \varphi^\dagger\right)$ and $\mathcal{S}\left(F\left(\varphi_0\right); g^\dagger\right)$ sufficiently small. Then the iterates $\left(\varphi_j\right)$ for exact data $g^{\mathrm{obs}} = g^\dagger$ fulfill

$$\Delta\left(\varphi_j, \varphi^\dagger\right) = \mathcal{O}\left(\Lambda^2\left(\alpha_j\right)\right),$$
$$\mathcal{S}\left(F\left(\varphi_j\right); g^\dagger\right) = \mathcal{O}\left(\Theta\left(\alpha_j\right)\right)$$

as $j \to \infty$, and in case of noisy data for sufficiently large $\tau \geq 1$ we get

$$\Delta\left(\varphi_{j_*}, \varphi^\dagger\right) = \mathcal{O}\left(\Lambda^2\left(\Theta^{-1}\left(\mathbf{err}_{j_*}\right)\right)\right) = \mathcal{O}\left(\frac{\mathbf{err}_{j_*}}{\Theta^{-1}\left(\mathbf{err}_{j_*}\right)}\right),$$
$$\mathcal{S}\left(F\left(\varphi_{j_*}\right); g^\dagger\right) = \mathcal{O}\left(\mathbf{err}_{j_*}\right).$$

Extensions

- Same convergence rates in terms of

$$\mathbf{err}_j := \frac{1}{C_{\mathrm{err}}} \mathbf{err}\left(F\left(\varphi_{j+1}\right)\right) + 2\eta C_{\mathrm{tc}} \mathbf{err}\left(F\left(\varphi_j\right)\right) + C_{\mathrm{tc}} C_{\mathrm{err}} \mathbf{err}\left(g^\dagger\right).$$

if the nonlinearity condition also holds for noisy data g^{obs}.

Extensions

- Same convergence rates in terms of

$$\mathbf{err}_j := \frac{1}{C_{\mathrm{err}}}\,\mathbf{err}\left(F\left(\varphi_{j+1}\right)\right) + 2\eta C_{\mathrm{tc}}\,\mathbf{err}\left(F\left(\varphi_j\right)\right) + C_{\mathrm{tc}} C_{\mathrm{err}}\,\mathbf{err}\left(g^\dagger\right).$$

  if the nonlinearity condition also holds for noisy data g^{obs}.

- Error decomposition and Lepskiĭ-type parameter choice rule in the case of an additive variational inequality

$$\left\langle \varphi_*^\dagger, \varphi^\dagger - \varphi \right\rangle \leq \beta_1 \Delta\left(\varphi, \varphi^\dagger\right) + \beta_2 \Lambda\left(\mathcal{S}\left(F\left(\varphi\right); g^\dagger\right)\right).$$

B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer.
A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators.
Inverse Problems, 23(3):987–1010, 2007.

Extensions

- Same convergence rates in terms of

$$\mathbf{err}_j := \frac{1}{C_{\mathrm{err}}}\,\mathbf{err}\left(F\left(\varphi_{j+1}\right)\right) + 2\eta C_{\mathrm{tc}}\,\mathbf{err}\left(F\left(\varphi_j\right)\right) + C_{\mathrm{tc}} C_{\mathrm{err}}\,\mathbf{err}\left(g^\dagger\right).$$

  if the nonlinearity condition also holds for noisy data g^{obs}.

- Error decomposition and Lepskiĭ-type parameter choice rule in the case of an additive variational inequality

$$\left\langle \varphi_*^\dagger, \varphi^\dagger - \varphi \right\rangle \leq \beta_1 \Delta\left(\varphi, \varphi^\dagger\right) + \beta_2 \Lambda\left(\mathcal{S}\left(F\left(\varphi\right); g^\dagger\right)\right).$$

B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer.
A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators.
Inverse Problems, 23(3):987–1010, 2007.

- Convergence rates under Hölder-type variational inequalities with index $\nu \in \left[\frac{1}{2}, 1\right)$ in combination with a Lipschitz assumption.

Outline

Poisson data

Let $\mathcal{Y} = L^1(\Omega, \nu) \cap L^\infty(\Omega, \nu)$ for some measure space (Ω, ν), and

$$F(\varphi) \geq 0 \quad \nu - \text{a.e.} \qquad \text{for all } \varphi \in \mathfrak{B}.$$

Moreover we assume that our noisy data g^{obs} fulfills $g^{\mathrm{obs}} \geq 0$, $g^{\mathrm{obs}} = 0$ where $g^\dagger = 0$ and

$$\int\limits_{\{g^\dagger > 0\}} \frac{|g^{\mathrm{obs}} - g^\dagger|^2}{g^\dagger} \, \mathrm{d}\nu \leq \frac{1}{t}$$

for some $t > 0$.

- This is motivated by the fact that for a Poisson process the variance decays like $\frac{1}{\sqrt{t}}$ where t is proportional to the expected number of photons.
- t can be interpreted as an illumination time and we want to study the limit $t \to \infty$.

Bounding **err**

The Kullback-Leibler divergence has a singularity at 0, so we define an offset version with $e > 0$ by

$$\mathcal{S}_e\left(g; g^{\mathrm{obs}}\right) = \int_\Omega g - \left(g^{\mathrm{obs}} + e\right) \ln\left(\frac{g+e}{e}\right) \,\mathrm{d}x$$

for $g \geq -\frac{e}{2}$. The deterministic noise model implies

$$\left|\mathcal{S}_e(g; g^{\mathrm{obs}}) - \mathcal{S}_e(g; g^\dagger)\right| \leq \sqrt{\frac{C}{t}}$$

for some constant $C > 0$ if $-\frac{e}{2} \leq g \leq B$. Hence the inequalities for $\mathbf{err}\left(F\left(\varphi\right) + F\left(\varphi; \psi - \varphi\right)\right),\ \varphi, \psi \in \mathfrak{B}$ are fulfilled with

$$C_{\mathrm{err}} = 1 \text{ and } \mathbf{err} \equiv \sqrt{\tfrac{C}{t}}.$$

Convergence theorem

Convergence rates

Let the Assumptions from above hold and assume that the nonlinearity condition is true for exact data. Moreover let

$$\sup_{\varphi,\psi\in\mathfrak{B}} \left\|F(\varphi) + F'(\varphi;\psi-\varphi)\right\|_{L^\infty} < \infty.$$

Then the a-priori stopping rule $j_* := \min\left\{j \in \mathbb{N} \;\middle|\; \Theta(\alpha_j) \leq \frac{\tau}{\sqrt{t}}\right\}$ with a sufficiently large parameter $\tau > 0$ leads to the convergence rates

$$\Delta\left(\varphi_{j_*}, \varphi^\dagger\right) = \mathcal{O}\left(\Lambda^2\left(\Theta^{-1}\left(t^{-1/2}\right)\right)\right),$$

$$\mathbb{KL}_e\left(F\left(\varphi_{j_*}\right); g^\dagger\right) = \mathcal{O}\left(t^{-1/2}\right).$$

Extensions

- If the nonlinearity condition also holds for noisy data g^{obs}, then the offset e can be set to 0 under a suitable variance condition on F.
  $\Rightarrow$ similar rates.

Extensions

- If the nonlinearity condition also holds for noisy data g^{obs}, then the offset e can be set to 0 under a suitable variance condition on F.
  $\Rightarrow$ similar rates.

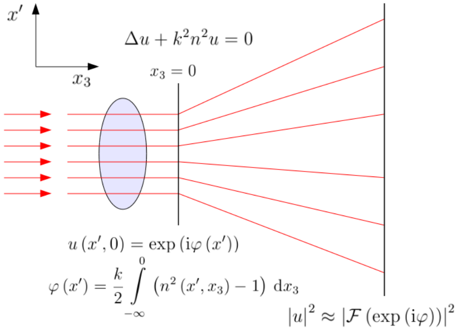
- Similar rates for a Lepskiĭ-type parameter choice rule in case of an additive variational inequality.

Extensions

- If the nonlinearity condition also holds for noisy data g^{obs}, then the offset e can be set to 0 under a suitable variance condition on F.
  $\Rightarrow$ similar rates.
- Similar rates for a Lepskiĭ-type parameter choice rule in case of an additive variational inequality.
- Ongoing work on convergence rates in case of full stochastic data, i.e. the data are given by a Poisson process.

Outline

The setting

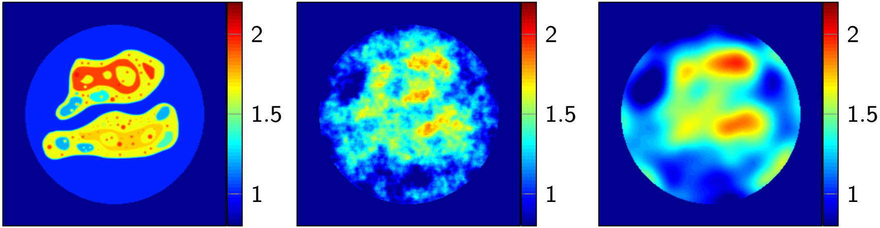


$$F : H^s(B_\rho) \longrightarrow L^\infty([-\kappa, \kappa]^2),$$

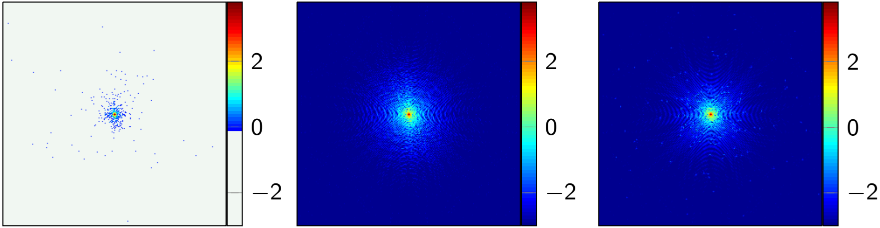
$$F(\varphi)(\xi) = \left| \int_{B_\rho} e^{-\mathrm{i}\xi \cdot x'} e^{\mathrm{i}\varphi(x')} \, \mathrm{d}x' \right|^2$$

M. V. Klibanov.
On the recovery of a 2-D function from the modulus of its Fourier transform.
J. Math. Anal. Appl., 323(2):818–843, 2006.

Median results for $t = 10^4$ expected counts



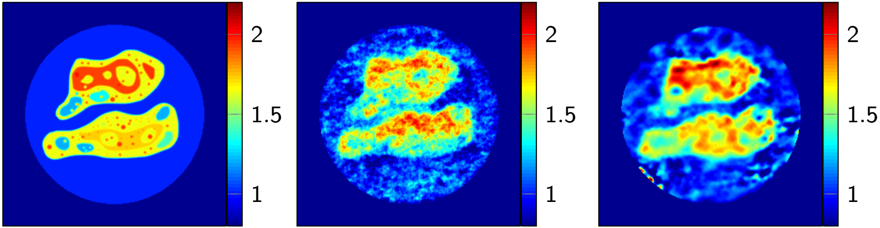
(a) exact solution (b) our algorithm (c) IRGNM



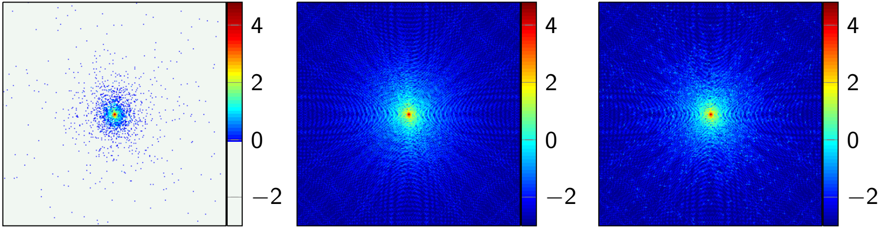
(d) observed data ($\log_{10}$) (e) exact data ($\log_{10}$) (f) our algorithm ($\log_{10}$)

K. Giewekemeyer et al, *Phys. Rev. A*, 83:023804, 2011.

Median results for $t = 10^5$ expected counts

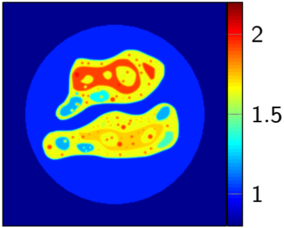


(a) exact solution (b) our algorithm (c) IRGNM

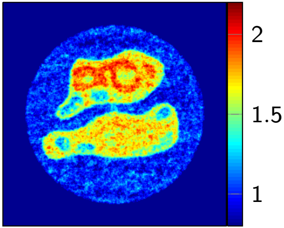


(d) observed data ($\log_{10}$) (e) exact data ($\log_{10}$) (f) our algorithm ($\log_{10}$)

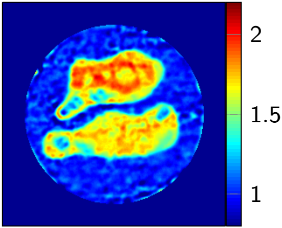
Median results for $t = 10^6$ expected counts



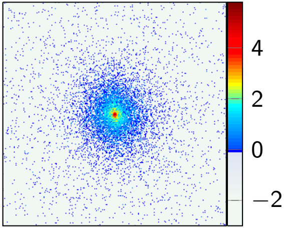
(a) exact solution



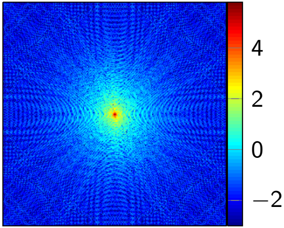
(b) our algorithm



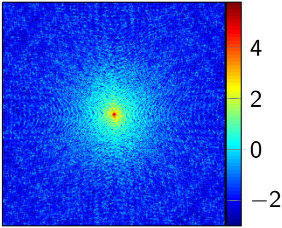
(c) IRGNM



(d) observed data ($\log_{10}$)



(e) exact data ($\log_{10}$)



(f) our algorithm ($\log_{10}$)

Outline

Shown results / Outlook

- Convergence analysis for iteratively regularized Newton methods with **arbitrary** data misfit functional and **arbitrary** penalty term.
- Our results include the known results for the IRGNM.
- Applications to Poisson data via choosing $\mathcal{S}$ to be the negative log-likelihood.
- Good numerical results in case of Poisson data.

T. Hohage and F. Werner.
Iteratively regularized Newton methods for general data misfit functionals and applications to Poisson data.
http://arxiv.org/abs/1105.2690v1, 2011.