Inverse problems with L^1 data fitting

Christian Clason, Bangti JIN, Karl Kunisch

Institute for Mathematics and Scientific Computing
Karl-Franzens-Universität Graz

Applied Inverse Problems 2009
Wien, July 24, 2009

Inverse problem

Find x such that

$$Kx = y^\delta$$

- $K : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ bounded linear operator
- $\Omega \subset \mathbb{R}^n$ bounded domain
- $y^\delta \in \mathrm{L}^2(\Omega)$ noisy measurement
- Impulsive noise (e.g., salt & pepper, random valued)

Inverse problem

L^1 data fitting

$$(\mathcal{P}) \qquad \min_{x \in L^2} \|Kx - y^\delta\|_{L^1} + \frac{\alpha}{2} \|x\|_{L^2}^2$$

- More robust in the presence of outliers
- Applicable in image processing, signal processing (single pixel failure)
- Challenge: Nondifferentiable functional, noise level unknown
- Regularization assumes smooth solution; alternative: TV

Fenchel duality

- V, Y Banach spaces, topological duals V^*, Y^*
- $\Lambda \in \mathcal{L}(V, Y)$, $\mathcal{F} : V \to \mathbb{R} \cup \{\infty\}$, $\mathcal{G} : Y \to \mathbb{R} \cup \{\infty\}$
- Fenchel conjugate of $\mathcal{F}$:

$$\mathcal{F}^* : V^* \to \mathbb{R} \cup \{\infty\}$$

$$\mathcal{F}^*(v^*) = \sup_{v \in V} \langle v^*, v \rangle_{V^*,V} - \mathcal{F}(v)$$

Fenchel duality

Fenchel duality theorem

- $\mathcal{F}, \mathcal{G}$ convex and lower semicontinuous
- $\exists v_0 \in V$: $\mathcal{F}(v_0) < \infty$, $\mathcal{G}(\Lambda v_0) < \infty$, $\mathcal{G}$ continuous at Λv_0:

(FD) $$\inf_{v \in V} \mathcal{F}(v) + \mathcal{G}(\Lambda v) = \sup_{q \in Y^*} -\mathcal{F}^*(\Lambda^* q) - \mathcal{G}^*(-q)$$

Extremality relations: $\overline{v}, \overline{q}$ solutions of (FD) iff

(ER) $$\begin{cases} \Lambda^* \overline{q} \in \partial \mathcal{F}(\overline{v}), \\ -\overline{q} \in \partial \mathcal{G}(\Lambda \overline{v}), \end{cases}$$

Dual problem

Define

$$\mathcal{F} : \mathrm{L}^2 \to \mathbb{R}, \qquad \mathcal{F}(v) = \frac{\alpha}{2} \|v\|_{\mathrm{L}^2}^2 ,$$
$$\mathcal{G} : \mathrm{L}^2 \to \mathbb{R}, \qquad \mathcal{G}(v) = \|v - y^\delta\|_{\mathrm{L}^1} ,$$
$$\Lambda : \mathrm{L}^2 \to \mathrm{L}^2, \qquad \Lambda v = K v.$$

Fenchel conjugates

$$\mathcal{F}^* : \mathrm{L}^2 \to \mathbb{R}, \qquad \mathcal{F}^*(q) = \frac{1}{2\alpha} \|q\|_{\mathrm{L}^2}^2 ,$$
$$\mathcal{G}^* : \mathrm{L}^2 \to \mathbb{R} \cup \{\infty\}, \qquad \mathcal{G}^*(q) = \begin{cases} \langle q, y^\delta \rangle_{\mathrm{L}^2} & \text{if } \|q\|_{\mathrm{L}^\infty} \leq 1, \\ \infty & \text{if } \|q\|_{\mathrm{L}^\infty} > 1. \end{cases}$$

Dual problem

Dual problem

$$(\mathcal{P}^*) \qquad \begin{cases} \min\limits_{p \in L^2} \dfrac{1}{2\alpha} \|K^* p\|_{L^2}^2 - \langle p, y^\delta \rangle_{L^2} \\ \text{s.t.} \quad \|p\|_{L^\infty} \leq 1, \end{cases}$$

- Fenchel duality theorem: $(\mathcal{P}^*)$ has solution p_α
- Solution not unique if $\ker K^* \neq \{0\}$!

Dual problem

Solutions x_α, p_α related by

$$\begin{cases} K^* p_\alpha = \alpha x_\alpha, \\ \qquad 0 \leq \langle K x_\alpha - y^\delta, p - p_\alpha \rangle_{\mathrm{L}^2}, \end{cases}$$

for all $p \in \mathrm{L}^2$ with $\|p\|_{\mathrm{L}^\infty} \leq 1$.

Given a solution p_α, unique solution of $(\mathcal{P})$:

$$x_\alpha = \frac{1}{\alpha} K^* p_\alpha$$

Characterization of minimizer

For all $p \in \mathrm{L}^2$, $p \geq 0$:

$$\begin{aligned} \langle Kx_\alpha - y^\delta, p\rangle_{\mathrm{L}^2} &= 0 \qquad \text{if } \operatorname{supp} p \subset \{x : |p_\alpha(x)| < 1\}, \\ \langle Kx_\alpha - y^\delta, p\rangle_{\mathrm{L}^2} &\geq 0 \qquad \text{if } \operatorname{supp} p \subset \{x : p_\alpha(x) = 1\}, \\ \langle Kx_\alpha - y^\delta, p\rangle_{\mathrm{L}^2} &\leq 0 \qquad \text{if } \operatorname{supp} p \subset \{x : p_\alpha(x) = -1\}. \end{aligned}$$

Interpretation:

- Box constraint on p_α active where data is not attained by x_α
- Sign of p_α gives sign of noise

Regularization of dual problem

$$(\mathcal{P}^*) \qquad \begin{cases} \min_{p \in \mathrm{L}^2} \frac{1}{2\alpha} \|K^* p\|_{\mathrm{L}^2}^2 - \langle p, y^\delta \rangle_{\mathrm{L}^2} \\ \text{s.t.} \quad \|p\|_{\mathrm{L}^\infty} \leq 1, \end{cases}$$

- Non-differentiable problem replaced by smooth box-constrained problem
- Moreau-Yosida regularization for $c > 0 \Rightarrow$ efficient solution by semismooth Newton method
- Superlinear convergence needs norm gap: Add smoothing term with $\beta > 0$

Regularization of dual problem

$$
(\mathcal{P}_c^*) \quad \begin{cases} \min_{p \in \mathrm{H}^1} \frac{1}{2\alpha} \|K^* p\|_{\mathrm{L}^2}^2 - \langle p, y^\delta \rangle_{\mathrm{L}^2} \\ + \frac{1}{2c} \|\max(0, c(p-1))\|_{\mathrm{L}^2}^2 + \frac{1}{2c} \|\min(0, c(p+1))\|_{\mathrm{L}^2}^2 \end{cases}
$$

- Non-differentiable problem replaced by smooth box-constrained problem
- Moreau-Yosida regularization for $c > 0 \Rightarrow$ efficient solution by semismooth Newton method
- Superlinear convergence needs norm gap: Add smoothing term with $\beta > 0$

Regularization of dual problem

$$(\mathcal{P}^*_{\beta,c}) \quad \begin{cases} \min_{p\in \mathrm{H}^1} \frac{1}{2\alpha} \|K^* p\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|\nabla p\|^2_{\mathrm{L}^2} - \langle p, y^\delta \rangle_{\mathrm{L}^2} \\ + \frac{1}{2c} \|\max(0, c(p-1))\|^2_{\mathrm{L}^2} + \frac{1}{2c} \|\min(0, c(p+1))\|^2_{\mathrm{L}^2} \end{cases}$$

- Non-differentiable problem replaced by smooth box-constrained problem
- Moreau-Yosida regularization for $c > 0 \Rightarrow$ efficient solution by semismooth Newton method
- Superlinear convergence needs norm gap: Add smoothing term with $\beta > 0$

Existence and convergence of minimizers

Regularized problem

$$\min_{p\in \mathrm{H}^1} \frac{1}{2\alpha}\|K^*p\|_{\mathrm{L}^2}^2 + \frac{\beta}{2}\|\nabla p\|_{\mathrm{L}^2}^2 - \langle p, y^\delta\rangle_{\mathrm{L}^2} + \frac{1}{2c}\|\max(0, c(p-1))\|_{\mathrm{L}^2}^2 + \frac{1}{2c}\|\min(0, c(p+1))\|_{\mathrm{L}^2}^2$$

Strictly convex if $\ker K^* \cap \ker \nabla = \{0\}$

$\Rightarrow$ Existence of unique minimizer $p_{\beta,c}$

Theorem (Convergence)

$$p_{\beta,c} \xrightarrow[c\to\infty]{\mathrm{H}^1} p_\beta \xrightarrow[\beta\to 0]{\mathrm{L}^2} p_\alpha, \qquad p_{\beta,c} \xrightarrow[\beta\to 0]{\mathrm{L}^2} p_c$$

Optimality system (regularized)

$$\begin{cases} \frac{1}{\alpha} K K^* p_c - \beta \Delta p_c - y^\delta + \lambda_c = 0, \\ \lambda_c = \max(0, c(p_c - 1)) + \min(0, c(p_c + 1)) \end{cases}$$

- Nonlinear equation for p_c
- Pointwise max, min semismooth
- $\Rightarrow$ solution by generalized Newton method

Semismoothness in function spaces

X, Y Banach spaces, $D \subset X$ open

Definition

$F : D \subset X \to Y$ **Newton differentiable** at $x \in D$, if there is neighborhood $N(x)$, $G : N(x) \to \mathcal{L}(X, Z)$

$$\|F(x+h) - F(x) - G(x+h)h\| = o(\|h\|)$$

Set $\{G(s) : s \in N(x)\}$ **Newton derivative** of F at x.

Definition

F **semismooth** if N-differentiable and $G(s)^{-1}$ uniformly bounded.

F semismooth $\Rightarrow$ generalized Newton method $G(s^k)\delta x = -F(x^k)$, $s^k \in N(x^k)$, converges locally superlinearly.

Semismoothness of projection operator

Projection operator

$$P(p) := \max(0, (p-1)) + \min(0, (p+1))$$

is semismooth from L^q to L^p, if and only if $q > p$,

Newton derivative

$$D_N P(p)h = h\chi_{\{|p|>1\}} := \begin{cases} h(x) & \text{if } |p(x)| > 1, \\ 0 & \text{if } |p(x)| \leq 1. \end{cases}$$

Application to optimality system

Can be written as $F(p) = 0$ with $F : \mathrm{H}^1 \to (\mathrm{H}^1)^*$,

$$F(p) := \frac{1}{\alpha} KK^* p - \beta \Delta p + \max(0, c(p-1)) + \min(0, c(p+1)) - y^\delta$$

Sobolev embedding, sum and chain rule for Newton derivatives
$\to$ F is semismooth, Newton derivative

Newton derivative

$$D_N F(p) h = \frac{1}{\alpha} KK^* h - \beta \Delta h + ch\chi_{\{|p|>1\}}$$

Computation of Newton step

Active sets

$$\mathcal{A}_k^+ := \{x : p^k(x) > 1\}$$
$$\mathcal{A}_k^- := \{x : p^k(x) < -1\}$$
$$\mathcal{A}_k := \mathcal{A}_k^+ \cup \mathcal{A}_k^-$$

Newton step

Given p^k, solve for p^{k+1}

$$\frac{1}{\alpha} KK^* p^{k+1} - \beta \Delta p^{k+1} + c\chi_{\mathcal{A}_k} p^{k+1} = y^\delta + c(\chi_{\mathcal{A}_k^+} - \chi_{\mathcal{A}_k^-})$$

Convergence of semismooth Newton method

Theorem

If $\|p_c - p^0\|_{\mathrm{H}^1}$ is sufficiently small, semismooth Newton method converges superlinearly in H^1 to $p_{\beta,c}$.

Theorem (Termination criterion)

$\mathcal{A}_{k+1}^+ = \mathcal{A}_k^+$ *and* $\mathcal{A}_{k+1}^- = \mathcal{A}_k^- \Rightarrow p^{k+1}$ *satisfies* $F(p^{k+1}) = 0$.

Choice of β

- Sufficiently large $\beta > 0$ necessary for numerical stability of Newton step

$$\left[\frac{1}{\alpha}KK^* - \beta\Delta + c\chi_{\mathcal{A}_k}\right] p^{k+1} = y^\delta + c(\chi_{\mathcal{A}_k^+} - \chi_{\mathcal{A}_k^-})$$

(KK^* ill-conditioned, $\chi_{\mathcal{A}_k}$ rank deficient)

- but introduces unwanted smoothing

$\Rightarrow$ optimal β: as small as possible!

Choice of β

Observation:

- System matrix well-conditioned $\Rightarrow$ solution $p_{\beta,c}$ feasible
- System matrix (numerically) singular $\Rightarrow \left\|p^{k+1}\right\|_\infty \approx c \gg 1$

Continuation strategy

- Choose $\beta_0 > 0$
- Compute $p_{\beta_n,c}$
- While $\|p_{\beta_n,c}\|_\infty \leq 1+\varepsilon$, set $\beta_{n+1} = \frac{1}{\tau}\beta_n$, continue minimization
- Else take last feasible iterate $p_{\beta_{N-1},c}$

Choice of α

x_α minimizer for fixed α:

Value function

$$F(\alpha) = \|Kx_\alpha - y^\delta\|_{L^1} + \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2$$

$F(\alpha)$ continuous, increasing, differentiable

Derivative

$$F'(\alpha) = \frac{1}{2}\|x_\alpha\|_{L^2}^2$$

Balancing principle

Idea:

Balance data residual (increasing in α)

$$\varphi(\alpha) = \|Kx_\alpha - y^\delta\|_{L^1}$$

with penalty term (decreasing in α)

$$\alpha F'(\alpha) = \frac{\alpha}{2}\|x\|_{L^2}^2$$

Choose α^* such that

$$(\sigma - 1)\varphi(\alpha^*) = \alpha^* F'(\alpha^*)$$

$\sigma > 1$ controls relative weight

Model function

Padé approximation of value function

$$m(\alpha) = b + \frac{c}{t + \alpha}.$$

c, t from interpolation conditions

$$m(\alpha) = F(\alpha), \quad m'(\alpha) = F'(\alpha),$$

b from asymptotic $\alpha \to \infty$: $F(\alpha) \to \|y^\delta\|_{L^1}$, $m(\alpha) \to b$

$$b = \|y^\delta\|_{L^1}$$

Fixed point iteration

1. Compute x_{α_k} (semismooth Newton method)
2. Evaluate $F(\alpha_k)$ and $F'(\alpha_k)$
3. Construct model function $m_k(\alpha) = b + \frac{c_k}{t_k+\alpha}$
4. Calculate intercept $\hat{m}$ of tangent of $m_k(\alpha)$ at $(\alpha_k, F(\alpha_k))$:

$$\hat{m} = F(\alpha_k) - \alpha_k F'(\alpha_k),$$

5. Solve for α_{k+1} in $m_k(\alpha_{k+1}) = \sigma \hat{m}$
6. repeat

Fixed point α^* satisfies

$$F(\alpha^*) = \sigma(F(\alpha^*) - \alpha^* F'(\alpha^*))$$

Fixed point iteration

Theorem (Existence)

For σ sufficiently close to 1 and $y^\delta \neq 0$, there exists at least one positive solution α^* to the balancing equation

$$(\sigma - 1)\varphi(\alpha) - \alpha F'(\alpha) = 0.$$

Theorem (Convergence)

If initial guess α_0 satisfies

$$(\sigma - 1)\varphi(\alpha_0) - \alpha_0 F'(\alpha_0) < 0,$$

fixed point iteration monotonically is decreasing, converges to α^*.

(Other inequality needs assumption on sign change close to α^*)

Noise estimate

Noise level satisfies

$$F(0) = \min_x \|Kx - y^\delta\|_{L^1} \leq \|Kx^* - y^\delta\|_{L^1}$$

Model function at final iteration k good approximation of F:

Noise level estimate

$$m_k(0) \approx \delta := \|y^\dagger - y^\delta\|_{L^1}$$

$y^\dagger$ exact solution

Numerical examples

- Examples from Regularization Tools
  `http://www2.imm.dtu.dk/~pch/Regutools/`
- Comparison with iteratively reweighted least squares (IRLS)
  Rodriguez, Wohlberg, IEEE Trans. Imag. Process., 18 (2009)
- Implementation in Matlab
- Code available:
  `http://www.uni-graz.at/~clason/codes/l1fitting.zip`

Noise

Impulsive random noise

$$y^\delta = \begin{cases} y^\dagger + \varepsilon\xi, & \text{with probability } r, \\ y, & \text{otherwise,} \end{cases}$$

- $y^\dagger$ exact data
- ξ normally distributed with mean 0 and standard deviation 1
- $r = 0.3$, $\varepsilon = 1$
- not salt & pepper!

Inverse heat conduction: `heat`

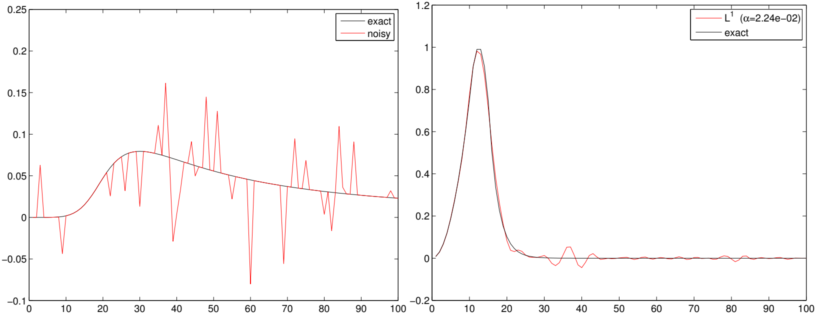
- K is Volterra integral operator of the first kind,

$$(Kx)(t) = \int_0^t k(s,t)x(s)\,ds,$$

$$k(s,t) = \frac{(s-t)^{-\frac{3}{2}}}{2\sqrt{\pi}} e^{-\frac{1}{4(s-t)}}$$

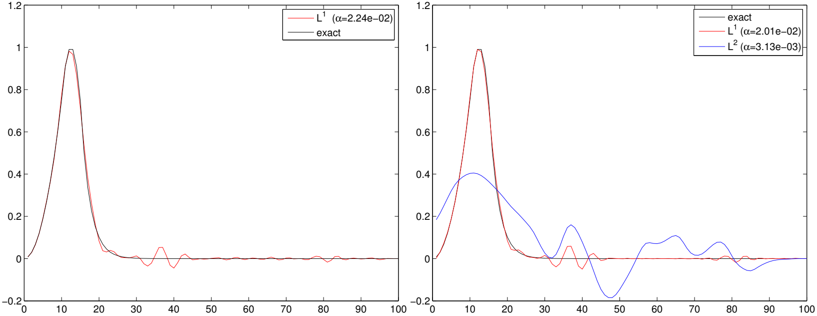
- Exponentially ill-posed, condition number $\approx 10^{37}$

heat: Results



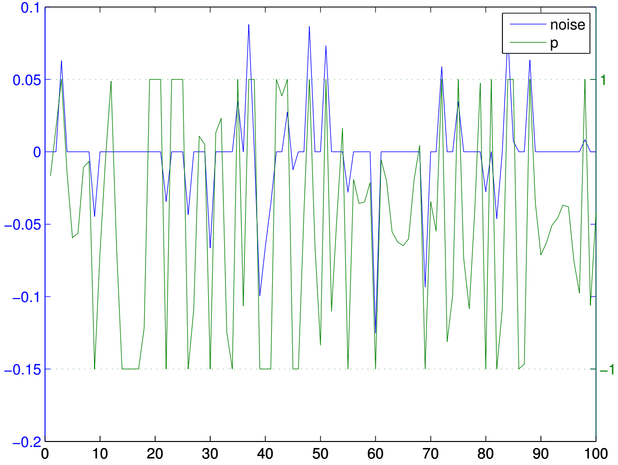
left: noisy data, right: adaptive α

heat: Results



left: adaptive α, right: optimal α and L^2 minimizer

`heat`: Results



noise and dual solution p

heat: Semismooth Newton method

Iterates in the path-following method for β:

β	iterations	e	$F(x)$	$\|\nabla p\|_{L^2}$
1.000e+0	2	2.248e-1	3.274e-2	2.951e-2
4.000e-2	2	1.855e-1	1.963e-2	1.090e-1
1.600e-3	2	1.610e-1	1.713e-2	2.191e-1
6.400e-5	2	1.556e-1	1.737e-2	1.408e+0
2.560e-6	6	2.250e-1	1.644e-2	5.852e+0
1.024e-7	4	7.042e-2	1.435e-2	6.436e+0
4.096e-9	3	2.043e-2	1.415e-2	6.902e+0
1.638e-10	10	1.563e-2	1.414e-2	7.361e+0
3.277e-11	10	1.546e-2	1.414e-2	8.960e+0

heat: Comparison with IRLS

- Reconstruction quality identical (hence not shown)
- Semismooth Newton method faster:

Computing time (in seconds) for the SSN vs. IRLS method

n	50	100	200	400	800	1600
t_{ssn} (sec)	0.011	0.034	0.188	1.163	8.052	39.07
t_{irls} (sec)	0.165	0.430	2.009	14.09	110.8	723.0

heat: Model function

Convergence in one iteration

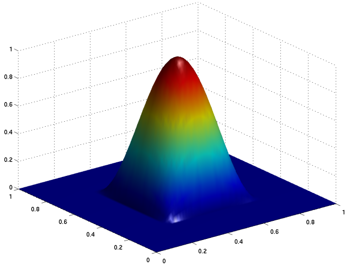
b: balancing principle, opt: optimal, e: error

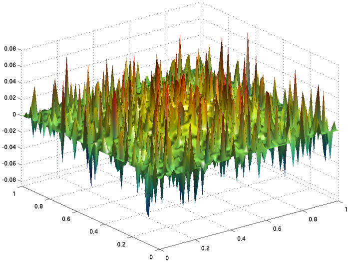
(r, ϵ)	δ	δ_b	α_b	α_{opt}	e_b	e_{opt}
(0.3,0.1)	1.390e-3	1.335e-3	1.402e-3	1.830e-2	1.860e-1	2.026e-2
(0.3,0.3)	4.170e-3	4.155e-3	6.638e-3	1.830e-2	4.515e-2	2.026e-2
(0.3,0.5)	6.950e-3	6.939e-3	1.135e-2	1.830e-2	2.706e-2	2.026e-2
(0.3,0.7)	9.731e-3	9.719e-3	1.604e-2	1.830e-2	2.103e-2	2.026e-2
(0.3,0.9)	1.251e-2	1.249e-2	2.083e-2	1.830e-2	2.037e-2	2.026e-2
(0.1,0.3)	7.438e-4	7.439e-4	1.227e-3	7.742e-4	1.727e-3	5.980e-4
(0.3,0.3)	4.170e-3	4.155e-3	6.638e-3	1.830e-2	4.515e-2	2.026e-2
(0.5,0.3)	7.871e-3	7.799e-3	1.225e-2	3.199e-2	5.635e-2	3.772e-2
(0.7,0.3)	1.110e-2	1.074e-2	2.254e-2	7.390e-2	1.118e-1	1.034e-1
(0.9,0.3)	1.570e-2	1.470e-2	2.247e-2	5.094e-2	1.662e-1	1.388e-1

Inverse source problem (2D)

- K is Δ^{-1}
- Discretization 64×64, $n = 4096$
- Mildly ill-posed, condition number $\approx 10^3$
- Convergence of fixed point iteration in 3 iterations, $\alpha = 8.797 \times 10^{-3}$
- Noise estimate 5.475×10^{-3} (exact: 5.490×10^{-3})

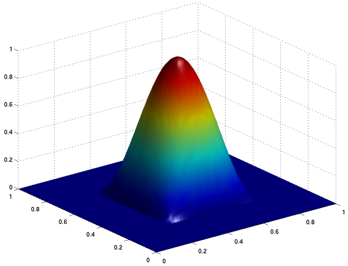
Inverse source problem: Results

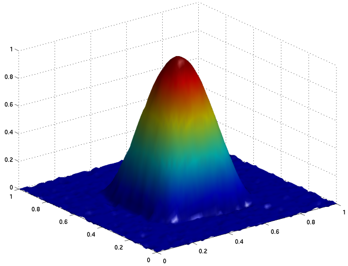




left: exact solution, right: noisy data

Inverse source problem: Results





left: exact solution, right: reconstruction

Conclusion

- L^1 data fitting term for impulsive noise
- Treatment of nondifferentiability via Fenchel duality and semismooth Newton method
- Automatic regularization parameter choice from balancing principle via model function and fixed point iteration
- Current work: Extension to $L^1 - TV$