Groupe hyperspectral

Part 2 - Inverse Problems in a Nutshell

José M. Bioucas Dias

Instituto de Telecomunicações
Instituto Superior Técnico
Universidade de Lisboa
Portugal

Société Française de Photogrammétrie et Télédétection
Grenoble, France, May, 2016

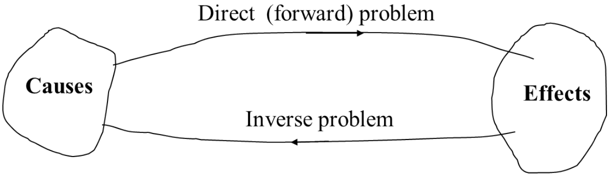
FCT, project UID/EEA/50008/2013

Outline

Part 2

- Direct (forward) and inverse problems
- Classes of direct problems. Examples
- Well-posed, ill-posed, and ill-conditioned inverse problems
- Curing ill-conditioned/ill-posed inverse problems
- The Bayesian philosophy. Bayesian estimators
- The regularization framework
- Widely used (convex) regularizers
- Proximity operators and proximity algorithms

Direct/Inverse problems



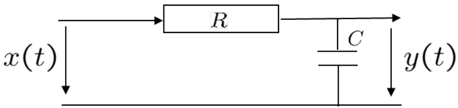
Example

Direct problem: the computation of the trajectories of bodies from the knowledge of the forces

Inverse problem: determination of the forces from the knowledge of the trajectories

Newton solved the first direct/inverse problem: the determintion of the gravitation force from the Kepler laws describing the trajectories of planets

Example: a linear time invariant (LTI) system



- Direct problem: $y(t) = x \star h(t) = \int x(t')h(t-t')dt'$

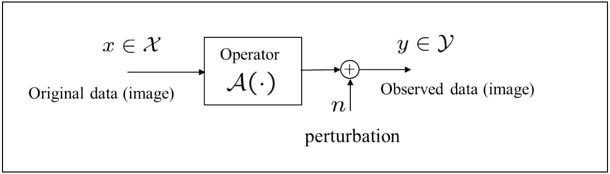
  Fourier domain: $\widetilde{y}(\omega) = \widetilde{h}(\omega)\widetilde{x}(\omega) \quad \Rightarrow \quad \widetilde{h}(\omega) = (1 + j\omega\tau)^{-1}, \quad \tau = RC$
- Inverse problem $\widetilde{x}(\omega) = \widetilde{y}(\omega)/h(\widetilde{\omega})$
- Source or dificulties: $\widetilde{h}^{-1}$ is unbounded: $|\widetilde{h}^{-1}(\omega)| \to \infty$ as $|\omega| \to \infty$

  $\Rightarrow$ A perturbation on $\widetilde{y}$ leads to a perturbation on $\widetilde{x}$ given by

$$\widetilde{\Delta} x(\omega) = \widetilde{\Delta} y(\omega)(1 + j\omega\tau)$$

high frequencies of the perturbation are amplified, degrading the estimate of $\widetilde{x}$

Classes of direct (forward) problems: linear operators

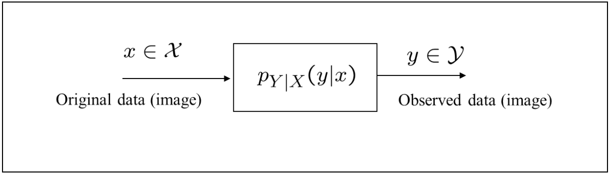


- Linear operators in Euclidean spaces

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad \mathbf{x} \in \mathbb{R}^n, \ \mathbf{y}, \mathbf{n} \in \mathbb{R}^m, \ \mathbf{A} \in \mathbb{R}^{m \times n}$$

- Applications:
  - image denoising, deconvolution, deblurring
  - X-ray tomography, MR imaging, radar imaging
  - imaging compressive sensing
  - image deblurring, superresolution, fusion
  - hyperspectral unmixing

Classes of direct problems: statistical observation models

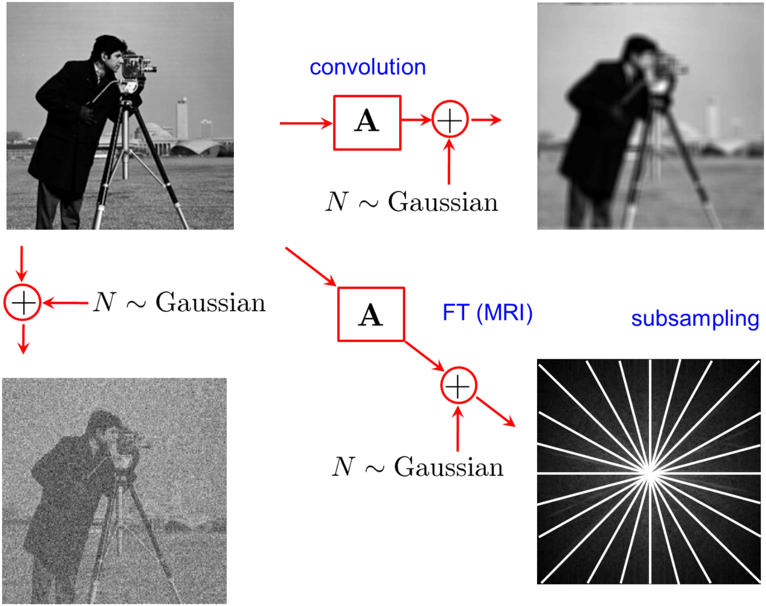


Examples: $(\mathbf{x} \in \mathbb{R}^n,\ \mathbf{y} \in \mathbb{R}^m,\ \mathbf{A} \in \mathbb{R}^{m\times n})$

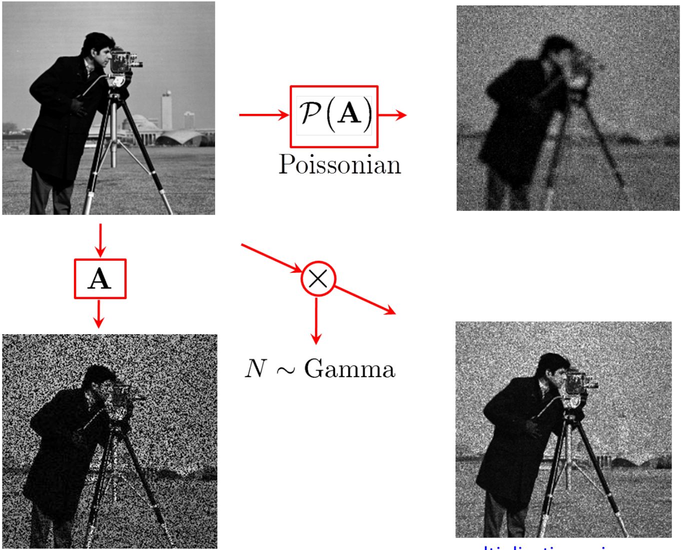
$$\mathcal{L}(\mathbf{A}\mathbf{x}) = -\log p_{Y|X}(\mathbf{y}|\mathbf{x}) \text{ where } \mathcal{L}(\mathbf{z}) \equiv \textstyle\sum_{i=1}^m \xi(z_i, y_i)$$

- Gaussian observations: $\xi_{\mathsf{G}}(z,y) = \frac{1}{2}(z-y)^2$
  (widely used in image restoration)
- Poissonian observations: $\xi_{\mathsf{P}}(z,y) = z + \iota_{\mathbb{R}_+}(z) - y\log(z_+)$
  (noise in photo-electric conversion, SPET (single photon emission tomography, PET (positron emission tomography))
- Multiplicative noise: $\xi_{\mathsf{M}}(z,y) = L(z + e^{y-z})$
  (radar, sonar)

Examples


convolution
A
$N \sim$ Gaussian
$N \sim$ Gaussian
A
FT (MRI)
subsampling
$N \sim$ Gaussian

Examples



subsampling

multiplicative noise

Example: hyperspectral pansharpening

$$\mathbf{y}_h = \mathbf{A}_h\mathbf{x} + \mathbf{n}_h \qquad \mathbf{y}_p = \mathbf{A}_p\mathbf{x} + \mathbf{n}_p$$

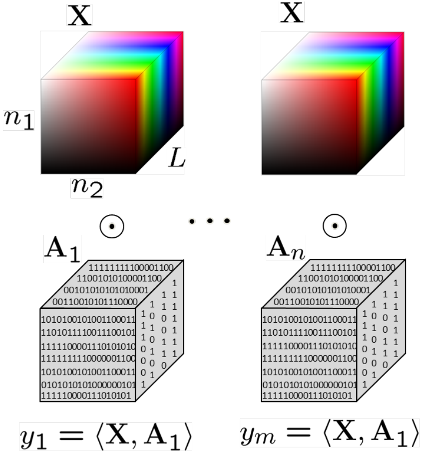
$\mathbf{x} \in \mathbb{R}^{600\times400\times100}$ $\quad$ $\mathbf{y}_h \in \mathbb{R}^{150\times100\times100}$ $\quad$ $\mathbf{y}_p \in \mathbb{R}^{600\times400\times1}$

original HS

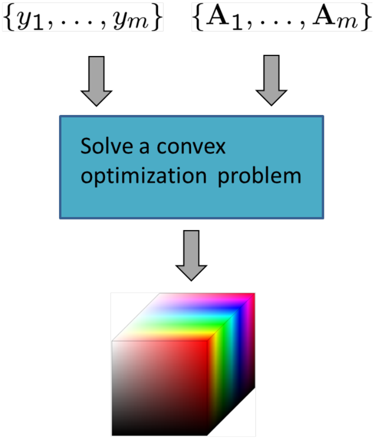
spatially blurred and downsampled HS

spectrally blurred HS

Example: hyperspectral compressive sensing



$$m \ll n_1 \times n_2 \times L$$



perfect reconstruction

Well-posed/ill-posed inverse problems [Hadamard, 1923]

Definition

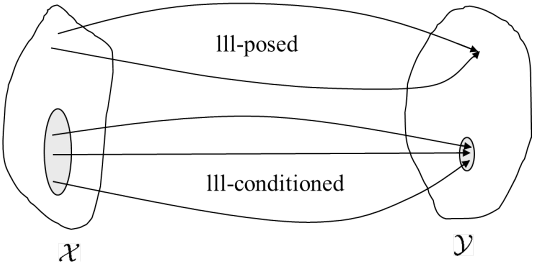
Let $\mathcal{A} : \mathcal{X} \rightarrow \mathcal{Y}$ be a (possible nonlinear) operator

The inverse problem of solving $\mathcal{A}(x) = y$ is well-posed in the Hadamard sense if:

1) A solution exists for any y in the observed data space
2) The solution is unique
3) The inverse mapping $y \mapsto x$ is continuous

- An inverse problem that is not well-posed is termed ill-posed
- The operator $\mathcal{A}$ of an inverse well/ill-posed problem is termed well/ill-posed

Ill-conditioned inverse problems



- Many well-posed inverse problems are ill-conditioned, in the sense that

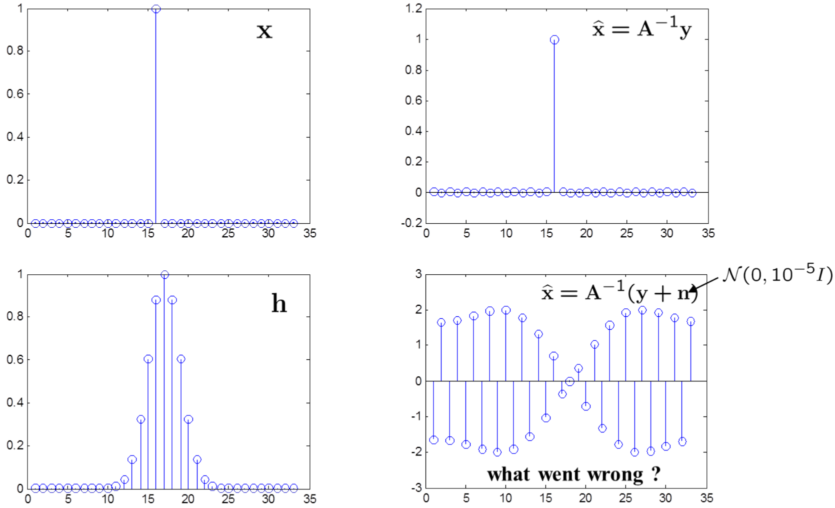
$$\frac{\|\Delta x\|}{\|x\|} \gg \frac{\|\Delta y\|}{\|y\|}$$

- for linear operators $\frac{\|\Delta x\|}{\|x\|} \leq \mathrm{cond}(\mathcal{A}) \frac{\|\Delta y\|}{\|y\|}$ (tight bound)

$$\mathrm{cond}(\mathcal{A}) \equiv \|A\| \, \|A^{-1}\|$$

An ill-conditioned IP: discrete deconvolution

Let $\mathbf{x}, \mathbf{y}, \mathbf{n} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{A} \in \mathbf{R}^{n \times n}$ represents a cyclic convolution with a Gaussian convolution kernel.



An ill-conditioned IP: 1D discrete deconvolution

Eigen-decomposition of cyclic matrices

- $\mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k \;\Rightarrow\; \mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^*$, $\quad \mathbf{V}^*\mathbf{V} = \mathbf{V}\mathbf{V}^* = \mathbf{I}$, $\quad$ (unitary)

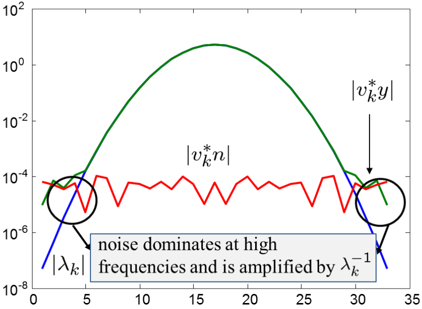
  $\mathbf{V} \rightarrow$ eigenvector (Fourier) matrix, $\boldsymbol{\Lambda} \rightarrow$ eigenvalue matrix (diagonal) matrix

- $\mathbf{A} = \sum_{k=1}^{n} \lambda_k \mathbf{v}_k \mathbf{v}_k^*$, $(\lambda_k \neq 0, k = 1, \ldots, n) \Rightarrow \mathbf{A}^{-1} = \displaystyle\sum_{k=1}^{n} \frac{1}{\lambda_k} \mathbf{v}_k \mathbf{v}_k^*$

- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$
- $\widehat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{y}$

  $=$

  $\mathbf{x} + \sum_{k=1}^{n} \frac{(\mathbf{v}_k^*\mathbf{n})}{\lambda_k} \mathbf{v}_k$



Example: 1D discrete deconvolution

- Regularization by filtering

$$\widehat{\mathbf{x}}_\alpha = \sum_{k=1}^{n} \lambda_k^{-1} w_\alpha(|\lambda_k|)\mathbf{v}_k(\mathbf{v}_k^* \mathbf{y})$$

such that

1. $w_\alpha(|\lambda|)\lambda^{-1} \to 0$ as $\lambda \to 0$
2. The larger eigenvalues are retained

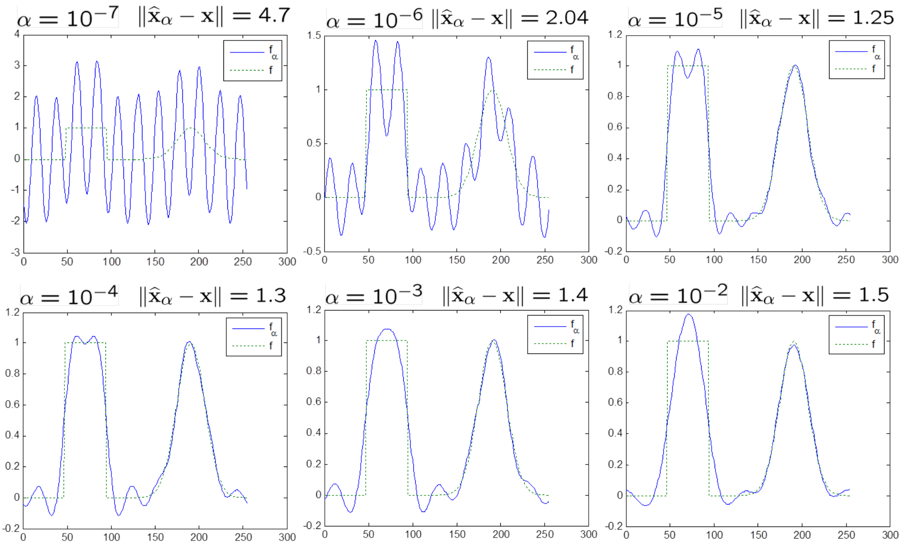
- Wiener filter

$$w_\alpha(|\lambda|) = \frac{|\lambda|^2}{|\lambda|^2 + \alpha} \qquad \widehat{\mathbf{x}}_\alpha = \sum_{k=1}^{n} \frac{\lambda_k^*}{|\lambda_k|^2 + \alpha} \mathbf{v}_k(\mathbf{v}_k^* \mathbf{y})$$

- Equivalent variational formulation (Tikhonov regularization)

$$\begin{aligned} \widehat{\mathbf{x}}_\alpha &= \arg\min_x \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_2^2 \\ &= (\mathbf{A}^*\mathbf{A} + \alpha\mathbf{I})^{-1}\mathbf{A}^*\mathbf{y} \end{aligned}$$

Example: 1D Discrete deconvolution



Example: 2D Discrete deconvolution

- convolution kernel h: uniform 9×9, $\quad \mathbf{n} \sim \mathcal{N}(0, \sigma^2 = 0.56^2)$
- $\text{cond}(\mathbf{A}) = 2.2 \times 10^5$, $\quad \text{SNR} = \dfrac{\text{var}(\mathbf{y})}{\sigma^2} = 40\text{dB}$
- $\mathbf{x}$ $\qquad\qquad$ $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$

Example: 2D Discrete deconvolution

$\alpha = 10^{-3}$ ISNR = -16dB

$\alpha = 5 \times 10^{-3}$ ISNR = -5dB

$\alpha = 3 \times 10^{-2}$ ISNR = 5.6dB

$\alpha = 10^{-1}$ ISNR = 3.2 dB

Curing Ill-posed/Ill-conditioned inverse problems

Golden rule for solving ill-posed/ill-conditioned inverse problems

Search for solutions which are:

1. compatible with the observed data
2. satisfy additional constraints (*a priori* or prior information) coming from the (physics) problem

Frameworks to solve inverse problems

- Bayesian inference: the causes are inferred by minimizing the Bayesian risk
- Regularization theory: the causes are inferred by minimizing a cost function

The Bayesian philosophy

Bayesian Inference

- Probabilities describe degrees of belief, not limiting relative frequencies
- We can make probability statements about parameters, even though they are fixed quantities
- Any inference should be based on the posterior

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$$

where $p_X(x)$ is the prior (or a priori) distribution that expresses our belief about x before we see any data and $p_Y(y)$ is the marginal on y

Bayesian estimators

Optimal Bayes estimator:

$$\widehat{x}_{\text{Bayes}} \in \arg\min_{\widehat{x}} \mathcal{R}(\widehat{x}|y)$$

where $\mathcal{R}(\widehat{x}|y)$ is the a posteriori expected loss

$$\mathcal{R}(\widehat{x}|y) = E[L(x,\widehat{x})|y] = \int L(x,\widehat{x})p(x|y)\,dx$$

and $L(x,\widehat{x})$ is a loss function that measures the discrepancy between x and $\widehat{x}$

- Zero-one loss function (Maximum Aposteriori Probability):

$$L_\varepsilon(x,\widehat{x}) = \begin{cases} 0 & \|x-\widehat{x}\|_2 \le \varepsilon \\ 1 & \|x-\widehat{x}\| > \varepsilon \end{cases} \Rightarrow \quad \begin{aligned} \widehat{x}_{\mathsf{MAP}} &= \arg\max_x p(x|y) \\ &= \arg\max_x p(y|x)p(x) \end{aligned}$$

- Quadratic loss(Posterior Mean):

$$L(x,\widehat{x}) = (x-\widehat{x})^T Q(x-\widehat{x}) \quad \Rightarrow \quad \widehat{x}_{\mathsf{PM}} = \mathbb{E}[x|y]$$

Regularization framework

- $f(\mathbf{x}, \mathbf{y}) \rightarrow$ data fidelity term: measures the compatibility between $\mathbf{x}$ and $\mathbf{y}$ (data-term, loss function, observation model, log-likelihood, ...)
- $\phi(\mathbf{x}) \rightarrow$ regularizer: expresses prior information about $\mathbf{x}$
- $\tau \rightarrow$ regularization parameter: sets the relative weight between the data term and the regularizer

Unconstrained and constrained formulations

1) Tickonov regularization: $\min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) + \tau\phi(\mathbf{x})$
2) Morozov regularization: $\min_{\mathbf{x}} \phi(\mathbf{x})$ subject to: $f(\mathbf{x}, \mathbf{y}) \leq \varepsilon$
3) Ivanov regularization: $\min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})$ subject to: $\phi(\mathbf{x}) \leq \delta$

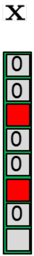
These formulations are equivalent under mild conditions [Lorenz & Worlicze, 13].
2) and 3) may take an unconstrained form using indicator functions

Widely used (convex) regularizers

Typical (sparseness-inducing) regularizer:

$$\phi(\mathbf{x}) = \|\mathbf{x}\|_1$$

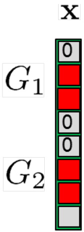
$\mathbf{x}$ contains representation coefficients (e.g., wavelet coeficients)



Typical frame-based analysis regularizer:

$$\phi(\mathbf{x}) = \|\mathbf{P}\mathbf{x}\|_1$$

$\mathbf{P}$ is an analysis operator



Group regularization (structured sparsity):

$$\phi(\mathbf{x}) = \sum_{i=1}^{k} \lambda_i \, \phi_i(\mathbf{x}_{G_i}), \quad G_i \subseteq \{1, ..., n\}$$

ϕ_i is the ℓ_1, ℓ_2, or ℓ_∞ norm (groups G_i may overlap)

Total variation regularization

Total variation regularization (promotes localized step gradients)
([Rudin, Osher, Fatemi, 92])

$$\mathsf{TV}_p(\mathbf{x}) = \sum_{i=1}^{n} \left(|x_i - x_{h_i}|^p + |x_i - x_{v_i}|^p\right)^{1/p}$$

where $h_i, v_i \in \{1, \ldots, n\}$ are the horizontal/vertical neighbors of i

Nonlocal total variation regularization (improved fine detail preservation)
([Osher et al., 05], [Elmoataz at al., 08])

$$\mathsf{NLTV}_p(\mathbf{x}) = \sum_{i=1}^{n} \left(\sum_{j \in \mathcal{N}_i} \omega_{i,j}^p |x_i - x_j|^p \right)^{1/p}$$

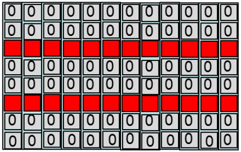
where $\mathcal{N}_i \subset \{1, \ldots, n\} \backslash \{i\}$

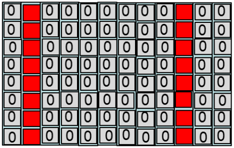
Example: TV and NLTV denoising

Regularization oriented to matrices

Mixed norms (structured sparseness-inducing) regularizers

$$\phi(\mathbf{X}) = \|\mathbf{X}\|_{p,1} = \sum_i \|\mathbf{x}^i\|_p \quad p \in \{2, \infty\}, \quad \phi(\mathbf{X}) = \|\mathbf{X}^T\|_{p,1}$$

$\mathbf{X}$ 

$\mathbf{X}$ 

Schatten p-norm ($p \geq 1$)

$$\|\mathbf{X}\|_{\mathcal{S}_p} = \left(\sum_i (\sigma_i(\mathbf{X}))^p\right)^{1/p}$$

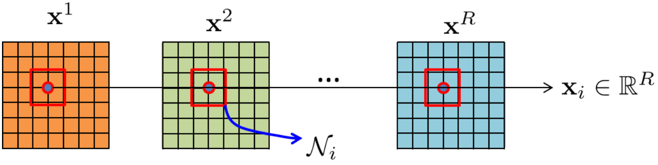
where $\sigma_i(\mathbf{X})$ is the ith singular value of $\mathbf{X}$.

$\|\mathbf{X}\|_{\mathcal{S}_1} \equiv \|\mathbf{X}\|_*$ (nuclear norm, promotes low rank)
$\|\mathbf{X}\|_{\mathcal{S}_2} \equiv \|\mathbf{X}\|_F$ (Frobenious norm)
$\|\mathbf{X}\|_{\mathcal{S}_\infty} = \sigma_1(\mathbf{X})$ (spectral norm)

Regularization oriented to hyperspectral (multiband) images

Let $\mathbf{X} \in \mathbb{R}^{R \times N}$ denote an hyperspectral image with R spectral bands and N pixels per band

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^R \end{bmatrix} \text{ (R band images)} \qquad \mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \text{ (N spectral vectors)}$$



$\mathbf{X}_i = [\mathbf{x}_i \, i \in \mathcal{N}_i]$

Structured tensor regularization

$$\phi_{\mathsf{ST}_p}(\mathbf{X}) = \sum_{i=1}^{N} \tau_i \|\mathbf{G}_i \mathbf{X}_i \mathbf{H}\|_{\mathcal{S}_p}$$

Regularization oriented to hyperspectral (multiband) images

Instances of structured tensor regularization

- ST-TV (other names: Vector TV, Hyperspectral TV) ([Bressom, Chan, 2008], [Yuan et al., 2012])

$$\phi_{\text{ST-TV}}(\mathbf{X}) = \sum_{i=1}^{N} \tau_i \|\mathbf{x}_i - \mathbf{x}_{h_i}, \mathbf{x}_i - \mathbf{x}_{v_i}\|_F$$

  ST-TV promotes localized step gradients within the image bands and aligns the "discontinuities" across the bands

- ST-NLTV$_p$ (other names: Multichannel-NLTV) ([Gilboa, Osher, 2009], [Cheng et al., 2009])

$$\phi_{\text{ST-NLTV}_p}(\mathbf{X}) = \sum_{i=1}^{N} \tau_i \big\| [(\omega_{i,j}(\mathbf{x}_i - \mathbf{x}_j), j \in \mathcal{N}_i] \big\|_{\mathcal{S}_p}$$

  ST-NLTV combines the characteristics of ST-TV and of NLTV

Proximal algorithms for solving convex inverse problems

Hyperspectral imaging inverse problems are challenging

- Hyperspectral imaging inverse problems are usually very large and often non-smooth
- Gradient-based algorithms cannot be used (e.g., nonlinear conjugate gradient or quasi-Newton)

Proximal algorithms: [Parikh et al., 13], [Komodakis & Pesquet, 14]

- new class of iterative methods suited to solve large scale non-smooth convex optimization problems
- replace a difficult problem with a sequence of simpler ones
- proximity operators, which may be interpreted as implicit subgradients, plays a central role in the proximal algorithms

Convex optimization and proximal algorithms

Proximity Operators ([Moreau 62], [Combettes, Wajs, 05], [Combettes, Pesquet, 07, 11], [Parikh, Boyd, 2013])

Moreau proximity operator (shrinkage/thresholding/denoising function

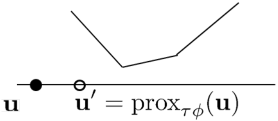
$$\mathrm{prox}_{\tau\phi}(\mathbf{u}) = \arg\min_{\mathbf{x}} \quad (1/2)\|\mathbf{x}-\mathbf{u}\|_2^2 + \tau\phi(\mathbf{x})$$

Projection onto a convex set

$$\iota_C(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in C \\ +\infty & \mathbf{x} \notin C \end{cases} \qquad \mathrm{prox}_{\tau\iota_C}(\mathbf{u}) = \arg\min_{\mathbf{x}\in C} \|\mathbf{x}-\mathbf{u}\|_2^2$$

Proximity operators generalize projections onto convex sets

Proximity operators have the flavor of gradient steps



Fixed points: $\mathbf{u}^*$ minimizes if and only if $\mathbf{u}^* = \mathrm{prox}_{\tau\phi}(\mathbf{u}^*)$

Moreau decomposition : $\mathbf{u} = \mathrm{prox}_{\phi}(\mathbf{u}) + \mathrm{prox}_{\phi^*}(\mathbf{u})$
(ϕ^* is the convex conjugate of ϕ)

Proximity Operators of widely used convex regularizers

ℓ_1-norm: $\phi(\mathbf{z}) = \|\mathbf{z}\|_1 \;\Rightarrow\; \mathrm{prox}_{\tau\phi}(\mathbf{u}) = \mathrm{soft}(\mathbf{u},\tau) := (|\mathbf{u}| - \tau)_+ \mathrm{sign}(\mathbf{u})$

ℓ_2-norm: $\phi(\mathbf{z}) = \|\mathbf{z}\|_2 \;\Rightarrow\; \mathrm{prox}_{\tau\phi}(\mathbf{u}) = \text{vect-soft}(\mathbf{u},\tau) := (\|\mathbf{u}\|_2 - \tau)_+ (\mathbf{u}/\|\mathbf{u}\|_2)$

ℓ_∞-norm: $\phi(\mathbf{z}) = \|\mathbf{z}\|_\infty \;\Rightarrow\; \mathrm{prox}_{\tau\phi}(\mathbf{u}) = \mathbf{u} - P_{B_{\ell_1(\tau)}}(\mathbf{u})$

---

nuclear-norm: $\phi(\mathbf{Z}) = \|\mathbf{Z}\|_* \;\Rightarrow\; \mathrm{prox}_{\tau\phi}(\mathbf{X}) = \mathbf{U}\mathcal{D}_\tau(\boldsymbol{\sigma})\mathbf{V}^{\mathbf{T}}$

$$[\mathbf{U},\Sigma,\mathbf{V}] = \mathrm{svd}(\mathbf{X}), \qquad \boldsymbol{\sigma} = \mathrm{diag}(\Sigma), \qquad \mathcal{D}_\tau(\Sigma) = \mathrm{diag}(\sigma_i - \tau)_+$$

Schatten p-norm $(p \geq 1)$: $\phi(\mathbf{Z}) = \|\mathbf{Z}\|_{\mathcal{S}_{\mathbf{p}}} \;\Rightarrow\; \mathrm{prox}_{\tau\phi}(\mathbf{X}) = \mathbf{U}\mathcal{D}_\tau(\boldsymbol{\Sigma})\mathbf{V}^{\mathbf{T}}$

$$[\mathbf{U},\Sigma,\mathbf{V}] = \mathrm{svd}(\mathbf{X}), \qquad \mathcal{D}_\tau(\Sigma) = \mathrm{diag}(\mathrm{prox}_{\tau\|\cdot\|_p}(\boldsymbol{\sigma}))$$

Proximal algorithms: SALSA ([Afonso, B-D, Figueiredo, 09, 10])

SALSA - (split augmented Lagrangian shrinkage algorithm) solves the optimization

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{j=1}^{J} g_j(\mathbf{H}^j\mathbf{x}), \qquad \begin{array}{l} \mathbf{x}\in\mathbb{R}^n \\ \mathbf{H}^j\in\mathbb{R}^{n_j\times n} \\ g_j : \mathbb{R}^{n_j}\mapsto\mathbb{R} \quad \text{convex, closed, proper} \end{array}$$

Algorithm 1: SALSA

initialization:

choose $(\mathbf{u}_0^j, \mathbf{d}_0^j) \in \mathbb{R}^{n_j\times n_j}$, $j = 1,\dots,J$

define $\mathbf{G} = \sum_{j=1}^{J}(\mathbf{H}^j)^T\mathbf{H}^j$

set $\mu \in]0,+\infty[$

for $k = 0, 1, \dots$ **do**

$\quad \mathbf{x}_{k+1} = \mathbf{G}^{-1}\left(\sum_{j=1}^{J}\mathbf{H}^j\left(\mathbf{u}_k^j + \mathbf{d}_k^j\right)\right)$

$\quad$ **for** $j = 1$ **to** J **do**

$\quad\quad \mathbf{u}_{k+1}^j = \text{prox}_{g_j/\mu}(\mathbf{H}^j\mathbf{x}_{k+1} - \mathbf{d}_k^j)$

$\quad\quad \mathbf{d}_{k+1}^j = \mathbf{d}_k^j - (\mathbf{H}^j\mathbf{x}_{k+1} - \mathbf{u}_{k+1}^j)$

return $\mathbf{x}_{k+1}$

Distinctive features

- Minimizes sums of convex terms
- The computation of proximity operators is parallelizable
- Offers flexibility in the choice of splittings
- Conditions for easy applicability:
  - inexpensive proximity operators
  - inexpensive matrix inversion
- Related algorithms: PPXA [Combettes, Pesquet, 08] SDMM [Setzer, et al., 10]

Proximal algorithms: FBPD ([Condat, 13], [Vũ,13])

FBPD - (Forward Backward Primal Dual) solves the optimization

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})+g(\mathbf{x})+h(\mathbf{G}\mathbf{x})$$

$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{G} \in \mathbb{R}^{m\times n}$

$f, h : \mathbb{R}^n \mapsto \mathbb{R}$ convex, closed, proper

$g : \mathbb{R}^n \mapsto \mathbb{R}$ convex, β-Lipschitz continuous gradient

Algorithm 2: FBPD

initialization:

choose $(\mathbf{x}_0, \mathbf{u}_0) \in \mathbb{R}^{n\times m}$

set $\tau > 0, \sigma > 0$ such that

$\tau(\beta/2 + \sigma\|G\|^2) < 1$

for $k = 0, 1, \ldots$ **do**

$\bar{\mathbf{x}}_k = \nabla g(\mathbf{x}_k - \mathbf{G}^T\mathbf{u}_k)$

$\mathbf{x}_{k+1} = \mathrm{prox}_{\tau f}(\mathbf{x}_k - \tau\bar{\mathbf{x}}_k)$

$\bar{\mathbf{u}}_k = \mathbf{G}(2\mathbf{x}_{k+1} - \mathbf{x}_k)$

$\mathbf{u}_{k+1} = \mathrm{prox}_{\sigma h^*}(\mathbf{u}_k + \sigma\bar{\mathbf{u}}_k)$

return $\mathbf{x}_{k+1}$

Distinctive features

- Minimizes sums of convex terms
- Does not involve matrix inversions
- Offers flexibility in the choice of splittings
- Conditions for easy applicability:
  - inexpensive proximity operators
- Related algorithms: FBF [Combettes & Pesquet, 12]