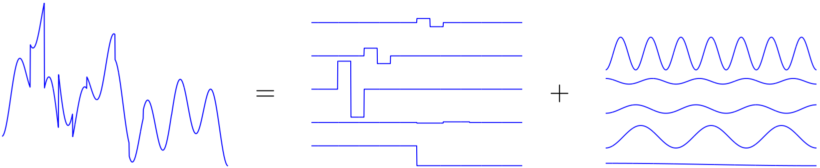
Algorithms for sparse optimization

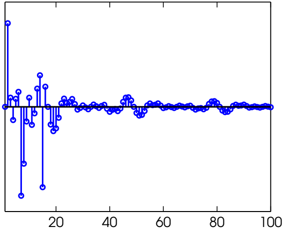
Michael P. Friedlander
University of California, Davis

Google
May 19, 2015

Sparse representations

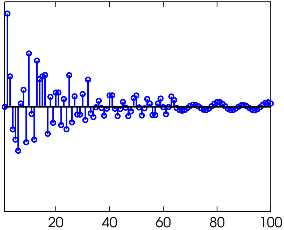


$y = Hx$



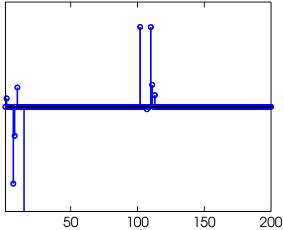
Haar

$y = Dx$



DCT

$y = \begin{bmatrix} H & D \end{bmatrix} x$



Haar/DCT dictionary

Sparsity via a transform

Represent a signal y as a superposition of elementary signal atoms:

$$y = \sum_j \phi_j x_j \qquad \text{ie,} \qquad y = \Phi x$$

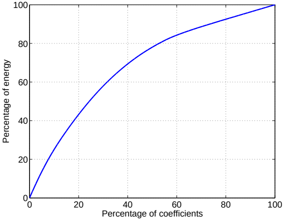
Orthogonal bases yield a unique representation:

$$x = \Phi^T y$$

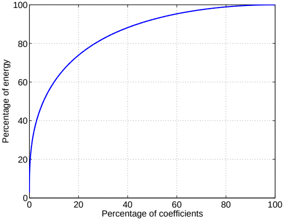
Overcomplete dictionaries, eg, $A = [\Phi_1 \ \Phi_2]$ have more flexibility.
But representation is not unique. One approach:

$$\text{minimize} \quad \text{nnz}(x) \quad \text{subj to} \quad Ax \approx y$$

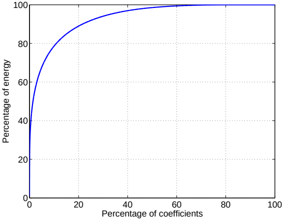
Efficient transforms



Identity

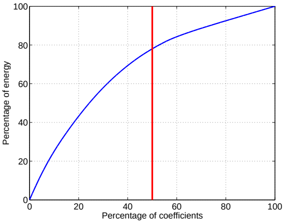


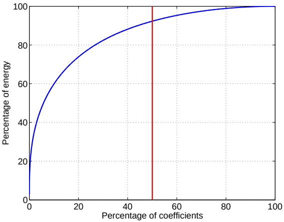
Fourier

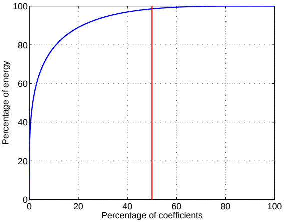


Wavelet

Efficient transforms





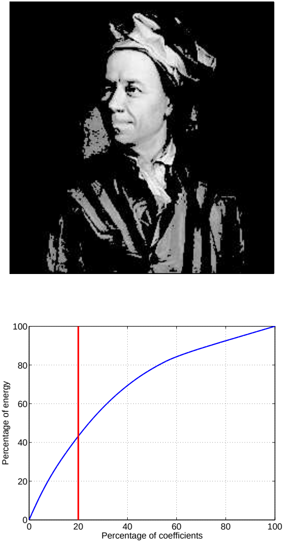


Identity

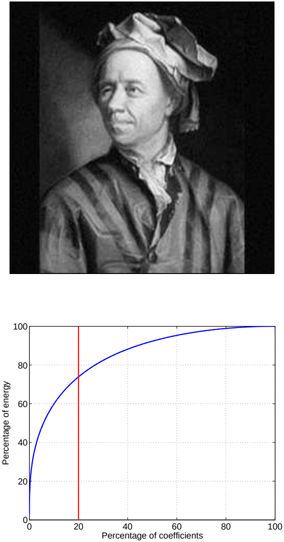
Fourier

Wavelet

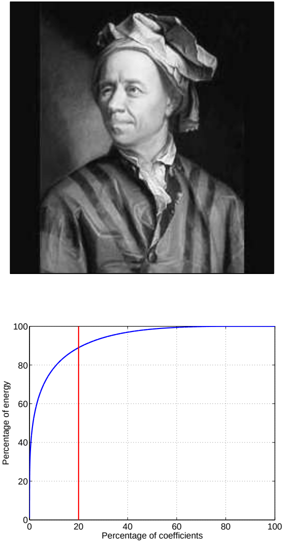
Efficient transforms



Identity

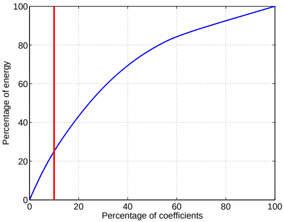


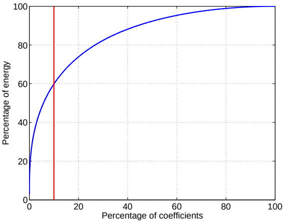
Fourier

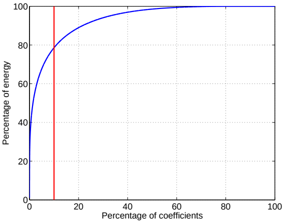


Wavelet

Efficient transforms





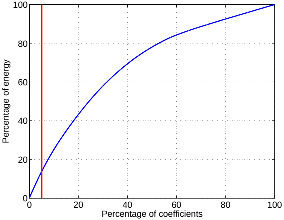


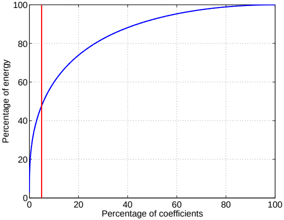
Identity

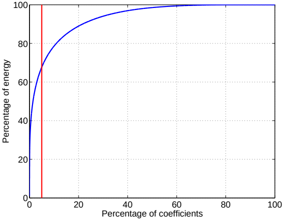
Fourier

Wavelet

Efficient transforms





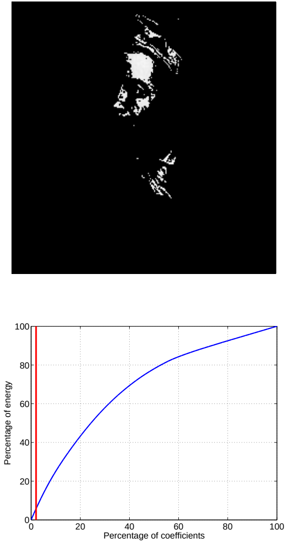


Identity

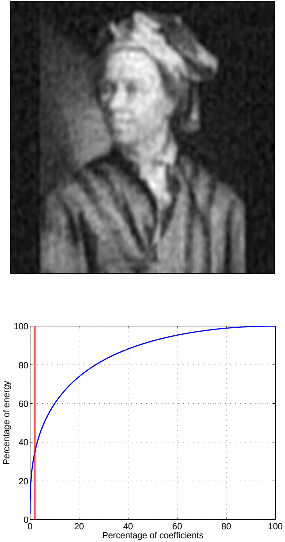
Fourier

Wavelet

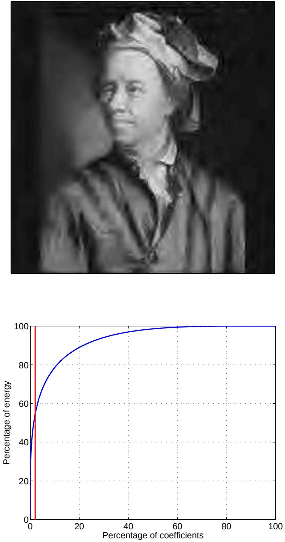
Efficient transforms



Identity



Fourier



Wavelet

Prototypical workflow

Measure structured signal y via linear measurements:

$$b_i = \langle f_i, y \rangle, \quad i = 1, \ldots, m$$

$$\text{ie,} \quad b = My, \quad M = \boxed{}$$

Decode the m measurements:

$$M\Phi x \approx b, \qquad x \text{ "sparse"}$$

Reconstruct the signal:

$$\widehat{y} := \Phi x$$

Guarantees: number of samples m needed to reconstruct signal w.h.p.

- Compressed sensing (eg, M Gaussian/Φ orthogonal): $\mathcal{O}(k \log \frac{n}{k})$
- matrix completion: # matrix samples $\mathcal{O}(n^{5/4} r \log n)$

Sparse optimization

Problem: find sparse solution x

$$\text{minimize} \quad \text{nnz}(x) \quad \text{subj to} \quad Ax \approx b$$

Convex relaxations, eg:

basis pursuit [Chen et al. (1998); Candès et al. (2006)]

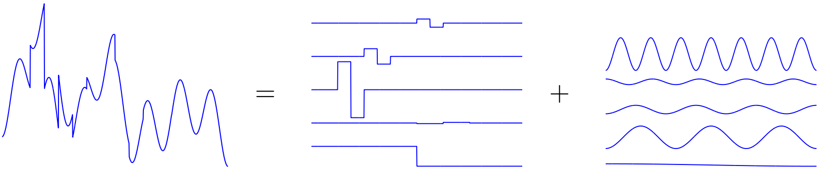
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|x\|_1 \qquad \text{subj to} \quad Ax \approx b \qquad (m \ll n)$$

nuclear norm [Fazel (2002); Candes and Recht (2009)]

$$\underset{X \in \mathbb{R}^{n \times p}}{\text{minimize}} \quad \textstyle\sum_i \sigma_i(X) \quad \text{subj to} \quad \mathcal{A}X \approx b \qquad (m \ll np)$$

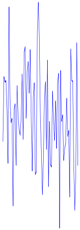
APPLICATIONS and FORMULATIONS

Compressed sensing



$\langle \phi_1, y\rangle, \ldots, \langle \phi_m, y\rangle$

$\boldsymbol{\Phi} y =$



$\boldsymbol{\Phi} = \boxed{\text{Gaussian}}$

$\boldsymbol{\Phi} y \approx \boldsymbol{\Phi} \boxed{\text{Haar}}\boxed{\text{DCT}}\, x \quad \Rightarrow \quad b \approx \boxed{A}\, x$

$$\text{minimize} \quad \|x\|_1 \quad \text{subj to} \quad Ax \approx b$$

Image deblurring

Observe $b = My$, y = image, M = blurring operator

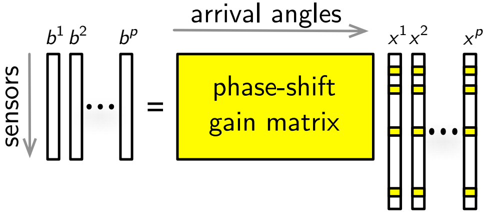
Recover significant coeff's via

$$\text{minimize} \quad \tfrac{1}{2}\|MWx - b\|^2 + \lambda\|x\|_1$$

λ	2.3e+0	2.3e-2	2.3e-4
$\|Ax - b\|$	2.6e+2	3.0e+0	3.1e-2
nnz(x)	72	1,741	28,484

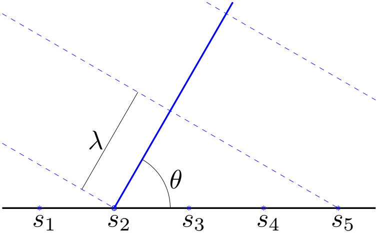
Sparco problem `blurrycam` [Figueiredo et al. (2007)]

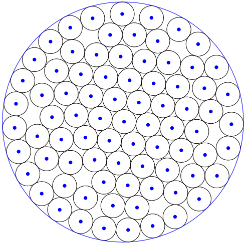
Joint sparsity: source localization

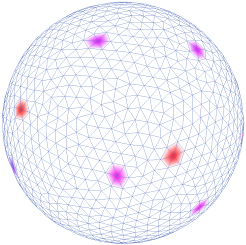


[Malioutov et al. (2005)]

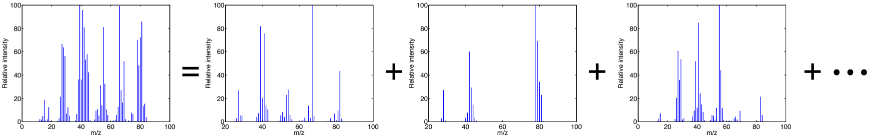
$$\begin{array}{ll} \text{minimize} & \|X\|_{1,2} \\ \text{subj to} & \|B - AX\|_F \le \sigma \end{array}$$



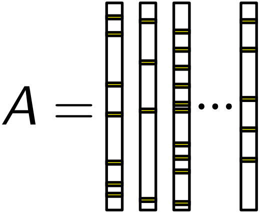




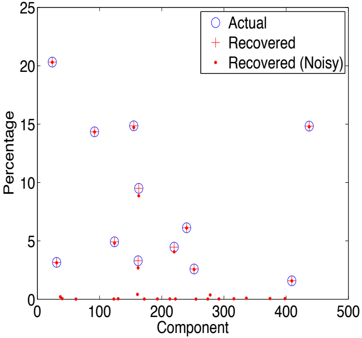
Mass spectrometry – separating mixtures



mixture　　molecule 1　　molecule 2　　molecule 3



spectral fingerprints



minimize $\sum_i x_i$ subj to $Ax \approx b,\ x \geq 0$ [Du-Angeletti '06, Berg-F. '11]

Phase retrieval

Observe signal $x \in \mathbb{C}^n$ through **magnitudes**.

$$b_k = |\langle a_k, x\rangle|^2, \qquad k = 1, \ldots, m$$

Lift.

$$b_k = |\langle a_k, x\rangle|^2 = \langle a_k a_k^*, xx^*\rangle = \langle a_k a_k^*, X\rangle \quad \text{with} \quad X = xx^*$$

Decode. Find X such that

$$\langle a_k a_k^*, X\rangle \approx b_k, \quad X \succeq 0, \quad \operatorname{rank}(X) = 1$$

Convex relaxation.

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \operatorname{tr}(X) \quad \text{subj to} \quad \langle a_k a_k^*, X\rangle \approx b_k, \quad X \succeq 0$$

[Chai et al. (2011); Candès et al. (2012); Waldspurger et al. (2015)]

Matrix completion

Observe random matrix entries of an $m \times p$ matrix Y.

$$b_k = Y_{i_k,j_k} \equiv e_{i_k}^T Y e_{j_k} = \langle e_{i_k} e_{j_k}^T, Y \rangle, \quad k = 1, \ldots, m$$

Decode. Find $m \times p$ matrix X such that

$$\langle e_{i_k} e_{j_k}^T, X \rangle \approx b_k, \quad \text{rank}(X) = \min$$

Convex relaxation.

$$\underset{X \in \mathbb{R}^{m \times p}}{\text{minimize}} \quad \sum_i \sigma_i(X) \quad \text{subj to} \quad \langle e_{i_k} e_{j_k}^T, X \rangle \approx b_k$$

[Fazel (2002); Candes and Recht (2009)]

CONVEX OPTIMIZATION

Convex optimization

$$\underset{x \in \mathcal{C}}{\text{minimize}} \quad f(x)$$

$f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $\qquad \mathcal{C} \subseteq \mathbb{R}^n$ is a convex set

Essential objective.

$$f(x) : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} = \begin{cases} f(x) & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise} \end{cases}$$

No loss of generality in restricting ourselves to unconstrained problems:

$$\underset{x}{\text{minimize}} \quad f(x) + g(Ax)$$

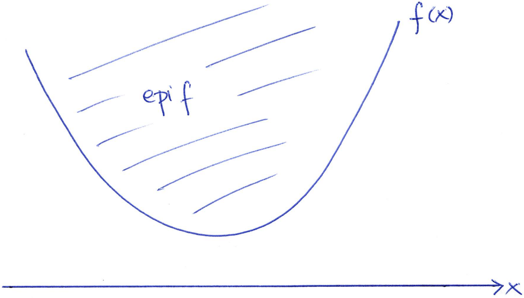
$$f : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad g : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad A \text{ is } m \times n \text{ matrix}$$

Convex functions

Definition. $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if its **epigraph**

$$\operatorname{epi} f = \{ (x, \tau) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \le \tau \}$$

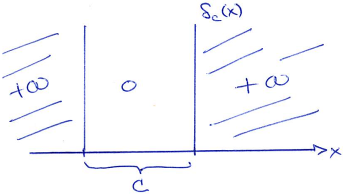
is convex.



Examples

Indicator on a convex set.

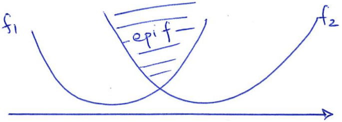
$$\delta_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{otherwise,} \end{cases} \qquad \operatorname{epi} \delta_{\mathcal{C}} = \mathcal{C} \times \mathbb{R}_+$$



Supremum of convex functions.

$$f := \sup_{j \in \mathcal{J}} f_j, \qquad \{f_j\}_{j \in \mathcal{J}} \text{ convex functions}$$

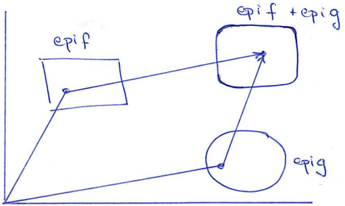
$$\operatorname{epi} f = \bigcap_{j \in \mathcal{J}} \operatorname{epi} f_j$$



Infimal convolution (epi-addition).

$$(f \,\square\, g)(x) = \inf_z \left\{ f(z) + g(x - z) \right\}$$

$$\operatorname{epi}(f \,\square\, g) = \operatorname{epi} f + \operatorname{epi} g$$



Examples

Generic form.

$$\underset{x}{\text{minimize}} \quad f(x) + g(Ax)$$

Basis pursuit.

$$\min_x \quad \|x\|_1 \quad \text{st} \quad Ax = b, \qquad\qquad f = \lambda\|\cdot\|_1, \qquad g = \delta_{\{b\}}$$

Basis pursuit denoising.

$$\min_x \quad \|x\|_1 \quad \text{st} \quad \|Ax - b\|_2 \le \sigma, \qquad f = \|\cdot\|_1, \qquad g = \delta_{\{\|\cdot - b\|_2 \le \sigma\}}$$

Lasso.

$$\min_x \quad \tfrac{1}{2}\|Ax - b\|^2 \quad \text{st} \quad \|x\|_1 \le \tau, \qquad f = \delta_{\tau\mathbb{B}_1}, \qquad g = \tfrac{1}{2}\|\cdot - b\|^2$$

DUALITY

Duality

$$\underset{x}{\text{minimize}} \quad f(x) + g(Ax)$$

Decouple the objective terms:

$$\underset{x,z}{\text{minimize}} \quad f(x) + g(z) \quad \text{subj to} \quad Ax = z$$

Lagrangian function.

$$L(x, z, y) := f(x) + g(z) + \langle y, Ax - z \rangle$$

encapsulates all the information about the problem:

$$\sup_y L(x, z, y) = \begin{cases} f(x) + g(z) & \text{if } Ax = z \\ +\infty & \text{otherwise} \end{cases}$$

$$p^* = \inf_{x,z} \sup_y L(x, z, y)$$

Primal-dual pair.

$$
\begin{aligned}
p^* &= \inf_{x,z} \sup_{y} \; L(x,z,y) \qquad &\text{(primal problem)} \\
d^* &= \sup_{y} \inf_{x,z} \; L(x,z,y) \qquad &\text{(dual problem)}
\end{aligned}
$$

The following **always** holds:

$$p^* \geq d^* \qquad \text{(weak duality)}$$

Under some **constraint qualification**,

$$p^* = d^* \qquad \text{(strong duality)}$$

Fenchel dual

Dual problem.

$$\sup_y \inf_{x,z} \; L(x,z,y) = \sup_y \inf_{x,z} \; f(x) + g(z) + \langle y, Ax - z\rangle$$
$$= \sup_y \left\{ -\sup_x \left[\langle A^T y, x\rangle - f(x)\right] - \sup_z \left[\langle -y, z\rangle - g(z)\right]\right\}$$

Conjugate function.

$$h^*(u) = \sup_x \{\langle u, x\rangle - h(u)\}$$

Fenchel-dual pair.

$$\underset{x}{\text{minimize}} \quad f(x) + g(Ax)$$
$$\underset{y}{\text{minimize}} \quad f^*(A^T y) + g^*(-y)$$

PROXIMAL METHODS

Proximal algorithm

$$\underset{x}{\text{minimize}} \quad f(x) \qquad (g \equiv 0)$$

Proximal operator.

$$\mathbf{prox}_{\alpha f}(x) := \arg\min_{z} \left\{ f(z) + \tfrac{1}{2\alpha}\|z - x\|^2 \right\} \quad \equiv \quad f \,\square\, \tfrac{1}{2\alpha}\|\cdot\|^2$$

Optimality. Every optimal solution x^* solves

$$x = \mathrm{prox}_{\alpha f}(x), \quad \forall \alpha > 0$$

Proximal iteration.

$$x^+ = \mathrm{prox}_{\alpha f}(x)$$

[Martinet (1970)]

$$x_{k+1} = \operatorname{prox}_{\alpha f}(x_k) := \arg\min_z \left\{ f(x) + \frac{1}{2\alpha}\|z - x\|^2 \right\}$$

Descent. Because x_{k+1} is the minimizing argument

$$f(x_{k+1}) + \frac{1}{2\alpha_k}\|x_{k+1} - x_k\|^2 \le f(x) + \frac{1}{2\alpha_k}\|x - x_k\|^2 \quad \forall x$$

Set $x = x_k$:

$$f(x_{k+1}) - f(x_k) \le -\frac{1}{2\alpha_k}\|x_{k+1} - x_k\|^2$$

Convergence.

- finite convergence for f polyhedral [Polyak and Tretyakov (1973)]
- subproblems may be solved approximately [Rockafellar (1976)]
- $f(x_k) - p^* \le \mathcal{O}(1/k)$ [Güler (1991)]
- $f(x_k) - p^* \le \mathcal{O}(1/k^2)$ via an accelerated variant [Güler (1992)]

Moreau-Yosida regularization

M-Y envelope $\quad f_\alpha(x) = \min_z \; f(z) + \frac{1}{2\alpha}\|z-x\|^2$

proximal map $\quad \mathrm{prox}_{\alpha f}(x) = \arg\min_z \; f(z) + \frac{1}{2\alpha}\|z-x\|^2$

Useful properties:

- differentiable: $\quad \nabla f_\alpha(x) = \frac{1}{\alpha}\left[x - \mathrm{prox}_{\alpha f}(x)\right]$
- preserves minima: $\quad \mathrm{argmin}\; f_\alpha = \mathrm{argmin}\; f$
- decomposition: $\quad x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x)$

Examples

vector 1-norm $\quad \mathrm{prox}_{\alpha\|\cdot\|_1}(x) = \mathrm{sgn}(x) \;.*\; \max\left\{|x| - \alpha, 0\right\}$

Schatten 1-norm $\quad \mathrm{prox}_{\alpha\|\cdot\|_1}(X) = U\bar{\Sigma}V^T, \qquad \bar{\Sigma} = \mathrm{Diag}\left[\mathrm{prox}_{\alpha\|\cdot\|_1}\left(\sigma(X)\right)\right]$

Augmented Lagrangian (AL) algorithm

Rockafellar ('73) connects AL to proximal method on dual

$$\boxed{\text{minimize } f(x) \quad \text{subj to} \quad Ax = b} \qquad \text{ie,} \qquad f(x) + \delta_{\{b\}}(Ax)$$

$$\text{Lagrangian:} \quad L\ (x, y) = f(x) + y^T(Ax - b)$$

$$\text{aug Lagrangian:} \quad L_\rho(x, y) = f(x) + y^T(Ax - b) + \tfrac{1}{2}\rho\|Ax - b\|^2$$

AL algorithm:

$$x^{k+1} = \arg\min_x \ L_\rho(x, y^k) \qquad \text{[may be inexact]}$$

$$y^{k+1} = y^k + \rho(Ax^{k+1} - b) \qquad \text{[multiplier update]}$$

- for linear constraints, AL algorithm $\equiv$ Bregman
- used by `L1-Bregman` for $f(x) = \|x\|_1$ [Yin et al. (2008)]
- proposed by Hestenes/Powell ('69) for smooth nonlinear optimization

SPLITTING METHODS

projected gradient, proximal-gradient, iterative soft-thresholding, alternating projections

Proximal gradient

$$\underset{x}{\text{minimize}} \quad f(x) + g(x)$$

Algorithm.

$$x^+ = \text{prox}_{\alpha g}\left(x - \alpha \nabla f(x)\right)$$

with

$$\alpha \in (\mu, 2/L), \quad \mu > 0, \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

Convergence.

- $f(x_k) - p^* \leq \mathcal{O}(1/k\ )$ constant stepsize ($\alpha \equiv 1/L$)
- $f(x_k) - p^* \leq \mathcal{O}(1/k^2)$ accelerated variant [Beck and Teboulle (2009)]
- $f(x_k) - p^* \leq \mathcal{O}(\gamma^k)$, $\gamma < 1$, with stronger assumptions

Interpretations

Majorization minimization. Lipshitz assumption on ∇f implies

$$f(z) \le f(x) + \langle \nabla f(x), z - x \rangle + \frac{L}{2}\|z - x\|^2 \quad \forall x, z$$

Proximal-gradient iteration minimizes the majorant: $\forall \alpha \in (0, 1/L)$

$$\begin{aligned} x^+ &= \mathrm{prox}_{\alpha g}\big(x - \alpha \nabla f(x)\big) \\ &\equiv \arg\min_z \Big\{ g(z) + f(x) + \langle \nabla f(x), z - x \rangle + \frac{1}{2\alpha}\|z - x\|^2 \Big\} \end{aligned}$$

Forward-backward splitting.

$$x^+ = \mathrm{prox}_{\alpha g}(x - \alpha \nabla f(x)) \;=\; \underbrace{(I + \alpha \partial g)^{-1}}_{\substack{\text{forward} \\ \text{step}}} \; \underbrace{(I - \alpha \nabla f)}_{\substack{\text{backward} \\ \text{step}}}(x)$$

Projected gradient

$\boxed{\text{minimize } f(x) \text{ subj to } x \in \mathcal{C}}$ $$x^{k+1} = \mathrm{proj}_{\mathcal{C}}(x^k - \alpha^k \nabla f(x^k))$$

LASSO

$$\underset{\|x\|_1 \le \tau}{\text{minimize}} \quad \tfrac{1}{2}\|Ax - b\|^2 \quad \text{via} \quad f(x) = \tfrac{1}{2}\|Ax - b\|^2, \quad g(x) = \delta_{\tau\mathbb{B}_1}(x)$$

- $\mathrm{proj}_{\mathcal{C}}(\cdot)$ can be computed in $\mathcal{O}(n)$ (randomized) [Duchi et al. (2008)]
- used by SPGL1 for Lasso subproblems [vdBerg and Friedlander (2008)]

BPDN (Lagrangian form)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1 \quad \text{via} \quad \underset{\bar{x} \in \mathbb{R}^{2n}_+}{\text{minimize}} \ \tfrac{1}{2}\|\bar{A}\bar{x} - b\|^2 + \lambda e^T\bar{x}$$

- $\mathrm{proj}_{\ge 0}(\cdot)$ can be computed in $\mathcal{O}(n)$
- GPSR uses Barzilai-Borwein for step α^k [Figueiredo-Nowak-Wright 07]

Iterative soft-thresholding (IST)

$$\boxed{\underset{x}{\text{minimize}} \;\; \tfrac{1}{2}\|Ax - b\|^2 \;+ \lambda\|x\|_1}$$

$$\begin{aligned} x^+ &= \text{prox}_{\alpha\lambda\|\cdot\|_1}(x - \alpha A^T r), \qquad r \equiv Ax - b \\ &= \mathcal{S}_{\alpha\lambda}(x - \alpha A^T r) \end{aligned}$$

where

$$\alpha \in (0,\, 2/\|A\|^2) \quad \text{and} \quad \mathcal{S}_\tau(x) = \text{sgn}(x) \cdot \max\{|x| - \tau,\, 0\}$$

[aka, iterative shrinkage, iterative Landweber]

Derived from various viewpoints:

- expectation-maximization [Figueiredo & Nowak '03]
- surrogate functionals [Daubechies et al '04]
- forward-backward splitting: [Combettes & Wajs '05]
- FISTA (accelerated variant) [Beck-Teboulle '09]

Low-rank approximation via nuclear-norm:

$$\boxed{\text{minimize} \quad \tfrac{1}{2}\|\mathcal{A}X - b\|^2 + \lambda\|X\|_n \quad \text{with} \quad \|X\|_n = \sum \sigma_j(X)}$$

Singular-value soft-thresholding algorithm:

$$x^+ = \mathcal{S}_{\alpha\lambda}(X - \alpha\mathcal{A}^*(r))$$

with

$$r := \mathcal{A}X - b, \qquad \mathcal{S}_\lambda(Z) = U\,\mathrm{diag}(\bar\sigma)V^T, \qquad \bar\sigma = \mathcal{S}_\lambda(\sigma(Z))$$

Computing $\mathcal{S}_\lambda$:

No need to compute full SVD of Z, only leading singular values/vectors

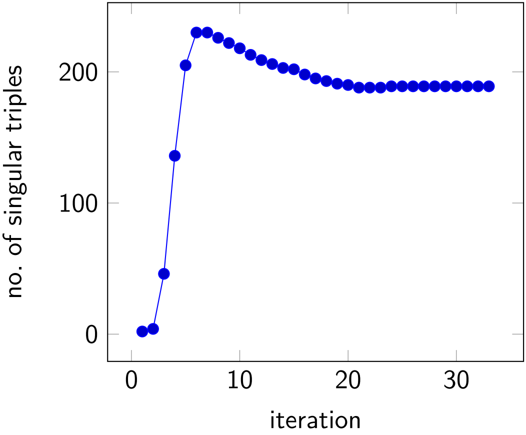
- Monte-Carlo approach [Ma-Goldfarb-Chen '08]
- Lanczos bidiagonalization (via PROPACK) [Cai-Candes-Shen '08; Jain et al '10]

Typical behavior

$$\underset{X \in \mathbb{R}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 \quad \text{subj to} \quad X_\Omega \approx b$$

recover 1000×1000 matrix X

$\text{rank}(X^*) = 4$



Backward-backward splitting

Approximate one of the functions via its (smooth) Moreau envelope.

$$\underset{x}{\text{minimize}} \quad f_\alpha(x) + g(x) \qquad \text{(both nonsmooth)}$$

$$f_\alpha = f \,\square\, \tfrac{1}{2\alpha}\|\cdot\|^2, \qquad \nabla f_\alpha(x) = \tfrac{1}{\alpha}\left[x - \mathrm{prox}_{\alpha f}(x)\right]$$

Proximal gradient.

$$x^+ = \mathrm{prox}_{\alpha g}\big(x - \alpha \nabla f_\alpha(x)\big) = \mathrm{prox}_{\alpha g}\ \mathrm{prox}_{\alpha f}(x)$$

Projection onto convex sets (POCS).

$$f = \delta_{\mathcal{C}}, \qquad g = \delta_{\mathcal{D}}$$

$$x^+ = \mathrm{proj}_{\mathcal{D}}\ \mathrm{proj}_{\mathcal{C}}(x)$$

solves the problem

$$\underset{x,z}{\text{minimize}} \quad \tfrac{1}{2}\|z - x\|^2 \quad \text{subj to} \quad x \in \mathcal{D},\ z \in \mathcal{C}$$

Alternating direction of method of multipliers (ADMM)

$$\underset{x}{\text{minimize}} \quad f(x) + g(x)$$

ADMM algorithm.

$$\begin{aligned} x^+ &= \text{prox}_{\alpha f}(z - u) \\ z^+ &= \text{prox}_{\alpha g}(x^+ + u) \\ u^+ &= u + x^+ - z^+ \end{aligned}$$

- used by `YALL1` [Zhang et al. (2011)]
- ADMM algo $\equiv$ Douglas-Rachford $\underset{\left(\substack{\text{linear}\\ \text{constraints}}\right)}{\equiv}$ split Bregman
- Esser's UCLA PhD thesis ('10) ; survey by Boyd et al. ('11)

PROXIMAL SMOOTHING

Primal smoothing

$$\underset{x}{\text{minimize}} \quad f(x) + g(Ax)$$

Partial smoothing via the Moreau envelope:

$$\underset{x}{\text{minimize}} \quad f_\mu(x) + g(Ax), \qquad f_\mu = f \,\square\, \tfrac{1}{2\mu}\|\cdot\|^2$$

Example (Basis Pursuit). Huber approximation to 1-norm:

$$f = \|\cdot\|_1, \quad g = \delta_{\{b\}} \quad \Longrightarrow \quad \underset{x}{\text{minimize}} \quad \|x\|_1 \quad \text{subj to} \quad Ax = b$$

Smoothed function is Huber with parameter μ:

$$f_\mu(x) = \begin{cases} x^2/2\mu & \text{if } |x| \le \mu \\ |x| - \mu/2 & \text{otherwise} \end{cases}$$

$$\mu \searrow 0 \quad \Longrightarrow \quad x_\mu^* \to x^*$$

Dual smoothing

Regularize the primal problem.

$$\underset{x}{\text{minimize}} \quad f(x) + \tfrac{\mu}{2}\|x\|^2 + g(Ax)$$

Dualize. Regularizing f smoothes the conjugate:

$$(f + \tfrac{\mu}{2}\|\cdot\|^2)^* = (f^* \,\square\, \tfrac{1}{2\mu}\|\cdot\|^2) = f^*_\mu$$

$$\underset{y}{\text{minimize}} \quad f^*_\mu(A^T y) + g^*(-y)$$

Example (Basis Pursuit).

$$f = \|\cdot\|_1, \qquad g = \delta_{\{b\}} \quad \Longrightarrow \quad \min_x \quad \|x\|_1 + \tfrac{\mu}{2}\|x\|^2 \quad \text{subj to} \quad Ax = b$$

$$f^*_\mu = \tfrac{1}{2\mu}\,\mathrm{dist}^2_{\mathbb{B}_\infty} \quad \Longrightarrow \quad \max_y \quad \langle b, y\rangle + \tfrac{1}{2\mu}\,\mathrm{dist}^2_{\mathbb{B}_\infty}(A^T y)$$

Exact regularization: $\quad x^*_\mu = x^*$ for all $\mu \in [0, \bar{\mu})$, $\quad \bar{\mu} > 0$

[Mangasarian and Meyer (1979); Friedlander and Tseng (2007)]

[Nesterov (2005); Beck and Teboulle (2012)]; TFOCS [Becker et al. (2011)]

CONCLUSION

much left unsaid, eg,

- optimal first-order methods [Nesterov (1988, 2005)]
- block coordinate descent [Sardy et al. (2004)]
- active-set methods (eg, LARS & Homotopy) [Osborne et al. (2000)]

my own work.

- avoiding proximal operators [Friedlander et al. (2014) (SIOPT)]
- greedy coordinate descent [Nutini et al. (2015) (ICML)]

email

- `mpfriedlander@gmail.com`

web

- `www.math.ucdavis.edu/~mpf`

REFERENCES

References I

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. **SIAM Journal on Imaging Sciences**, 2(1):183–202, 2009.

A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. **SIAM J. Optim.**, 22(2):557–580, 2012.

S. Becker, E. Candès, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. **Math. Program. Comp.**, 3:165–218, 2011. doi: 10.1007/s12532-011-0029-5.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. **Foundations and Trends in Machine Learning**, 3(1):1–122, 2011.

J.-F. Cai, E. J. Candés, and Z. Shen. A singular value thresholding algorithm for matrix completion. **SIAM J. Optim.**, 4(20):1956–1982, 2010. doi: 10.1137/080738970.

E. J. Candes and B. Recht. Exact matrix completion via convex optimization. **Found. Comput. Math.**, 9:717–772, December 2009.

References II

E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. **Comm. Pure Appl. Math.**, 59(8): 1207–1223, 2006.

E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. **Commun. Pur. Appl. Ana.**, 2012.

A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. **Inverse Problems**, 27(1):015005, 2011. URL `http://stacks.iop.org/0266-5611/27/i=1/a=015005`.

S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. **SIAM J. Sci. Comput.**, 20(1):33–61, 1998.

P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. **Multiscale Model. Simul.**, 4(4):1168–1200, 2005.

I. Daubechies, M. Defrise, and C. D. Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. **Comm. Pure Appl. Math.**, 57:1413–1457, 2004.

J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In **Proceedings of the 25th International Conference on Machine Learning**, pages 272–279, 2008.

References III

J. E. Esser. **Primal Dual Algorithms for Convex Models and Applications to Image Restoration, Registration and Nonlocal Inpainting**. PhD thesis, University of California, Los Angeles, 2010.

M. Fazel. **Matrix rank minimization with applications**. PhD thesis, Elec. Eng. Dept, Stanford University, 2002.

M. Figueiredo, R. Nowak, and S. J. Wright. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. **Sel. Top. in Signal Process., IEEE J.**, 1(4):586–597, 2007.

M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. **IEEE Trans. Image Process.**, 12(8):906–916, August 2003.

M. P. Friedlander and P. Tseng. Exact regularization of convex programs. **SIAM J. Optim.**, 18(4):1326–1350, 2007. doi: 10.1137/060675320.

M. P. Friedlander, I. Macêdo, and T. K. Pong. Gauge optimization and duality. **SIAM J. Optim.**, 24(4):1999–2022, 2014. doi: 10.1137/130940785.

O. Güler. On the convergence of the proximal point algorithm for convex minimization. **SIAM Journal on Control and Optimization**, 29(2):403–419, 1991. doi: 10.1137/0329022. URL `http://dx.doi.org/10.1137/0329022`.

References IV

O. Güler. New proximal point algorithms for convex minimization. **SIAM Journal on Optimization**, 2(4):649–664, 1992. doi: 10.1137/0802032. URL `http://dx.doi.org/10.1137/0802032`.

M. R. Hestenes. Multiplier and gradient methods. **J. Optim. Theory Appl.**, 4 (5):303–320, 1969.

S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. **Math. Program.**, 128:321–353, 2011. doi: 10.1007/s10107-009-0306-5.

D. Malioutov, M. Çetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. **IEEE Trans. Sig. Proc.**, 53(8):3010–3022, August 2005.

O. L. Mangasarian and R. R. Meyer. Nonlinear perturbation of linear programs. **SIAM J. Control Optim.**, 17(6):745–752, November 1979.

B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. **Rev. FranÂÿcaise Informat. Recherche OpÂťerationnelle**, 4 (Ser. R-3):154âĂŞ158, 1970.

Y. Nesterov. On an approach to the construction of optimal methods of minimization of smooth convex functions. **Ekonom. i. Mat. Metody**, 24, 1988.

References V

Y. Nesterov. Smooth minimization of non-smooth functions. **Math. Program.**, 103:127–152, 2005.

P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, **Advances in Neural Information Processing Systems 26**, pages 2796–2804. Curran Associates, Inc., 2013. URL `http://papers.nips.cc/paper/5041-pha`.

J. Nutini, I. Laradji, M. Schmidt, and M. Friedlander. Coordinate descent converges faster with the gauss-southwell rule than random selection. In **Inter. Conf. Mach. Learning**, 2015.

M. R. Osborne, B. Presnell, and B. A. Turlach. A new approach to variable selection in least squares problems. **IMA J. Numer. Anal.**, 20(3):389–403, 2000.

B. Polyak and N. Tretyakov. The method of penalty bounds for constrained extremum problems. **Zh. Vych. Mat i Mat. Fiz**, 13:34–46, 1973.

M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, **Optimization**, chapter 19. Academic Press, London and New York, 1969.

References VI

R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. **Mathematics of Operations Research**, 1(2):97–116, 1976.

S. Sardy, A. Antoniadis, and P. Tseng. Automatic smoothing with wavelets for a wide class of distributions. **J. Comput. Graph. Stat.**, 13, 2004.

E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. **SIAM J. Optim.**, 31(2):890–912, 2008. doi: 10.1137/080714488.

I. Waldspurger, A. dâĂŹAspremont, and S. Mallat. Phase recovery, maxcut and complex semidefinite programming. **Math. Program.**, 149(1-2):47–81, 2015.

W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for l1 minimization with applications to compressed sensing. **SIAM J. Imag. Sci.**, 1, 2008.

Y. Zhang, W. Deng, J. Yang, , and W. Yin, 2011. URL http://yall1.blogs.rice.edu/.