Optimization: why?

We think that convex optimization is an important enough topic that everyone who uses computational mathematics should know at least a little bit about it. In our opinion, convex optimization is a natural next topic after advanced linear algebra and linear programming.

(Stephen Boyd and Lieven Vandenberghe)

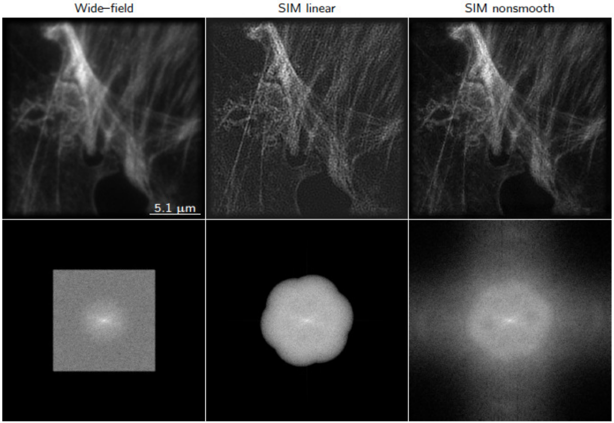
Optimization: when ?

Optimization problems arise naturally in many application fields. Whatever people do, at some point they get a craving to organize things in a best possible way. This intention, converted in a mathematical form, turns out to be an optimization problem of certain type.

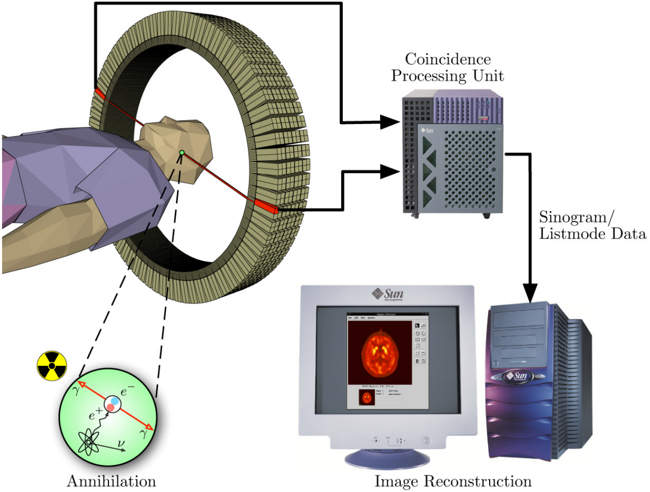
(Yurii Nesterov)

Structured illumination microscopy

$$\underset{\mathbf{u}}{\text{minimize}} \ \frac{1}{2}\|A\mathbf{u} - \mathbf{g}\|_2^2 + \lambda\|\Gamma\mathbf{u}\|_p^p$$

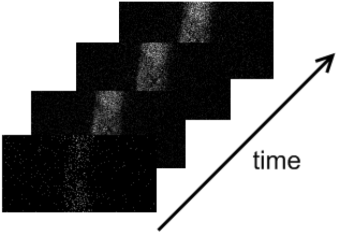


Positron emission tomography

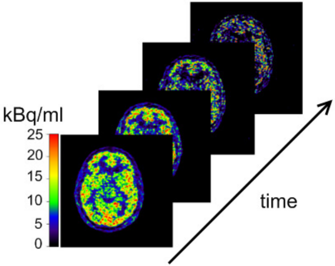


Positron emission tomography

$$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|A\mathbf{u}-\mathbf{g}\|_2^2+\lambda\|\Gamma\mathbf{u}\|_p^p$$



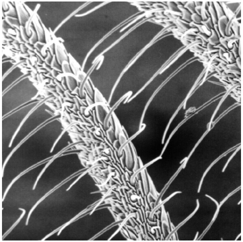
Measurements



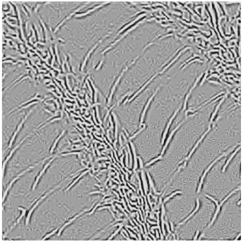
Reconstructed images

Cartoon-texture decomposition

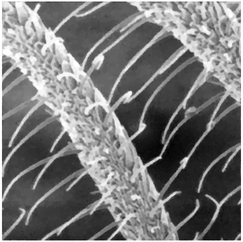
$$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}+\mathbf{v}-\mathbf{g}\|_2^2+\lambda\mathrm{TV}(\mathbf{u})+\eta\|\Gamma\mathbf{v}\|_p^p$$



Observations
$\mathbf{g}$



Texture
$\widehat{\mathbf{v}}$



Cartoon
$\widehat{\mathbf{u}}$

Robust PCA

$$\underset{u,v}{\text{minimize}} \|\mathbf{u}\|_* + \|\mathbf{v}\|_1 \quad \text{s.t.} \quad \mathbf{g} = \mathbf{u} + \mathbf{v}$$

[From Goldfarb, Ma, Sheinberg, 2010]

Mumford-Shah

$$\underset{\mathbf{u},K}{\text{minimize}} \; \underbrace{\frac{1}{2}\int_{\Omega}(\mathbf{u}-\mathbf{g})^2 dxdy}_{\text{fidelity}} + \underbrace{\beta \int_{\Omega\setminus K} |\nabla \mathbf{u}|^2 dxdy}_{\text{smoothness}} + \underbrace{\lambda |K|}_{\text{length}}$$

[Mumford-Shah, 1989]

$\mathbf{g}$

$\widehat{\mathbf{u}}$

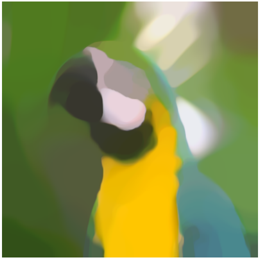
$\widehat{\mathbf{e}}$

Total variation model

$$\underset{\mathbf{u}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{g}\|_2^2 + \lambda \mathrm{TV}(\mathbf{u})$$

$\mathbf{g}$

$\widehat{\mathbf{u}}$ with $\lambda = 100$



$\widehat{\mathbf{u}}$ with $\lambda = 500$

Chan-Vese model

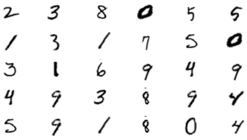
$$\underset{\Omega^{(1)},\dots,\Omega^{(Q)}}{\text{minimize}} \sum_{q=1}^{Q} \sum_{n\in\Omega^{(q)}} \eta_n^{(q)} + \lambda \sum_{q=1}^{Q} \mathrm{Per}(\Omega^{(q)}) \quad \text{s.t.} \quad \begin{cases} \bigcup_{q=1}^{Q} \Omega^{(q)} = \Omega, \\ (\forall q \neq p),\ \Omega^{(q)} \cap \Omega^{(p)} \end{cases}$$

g

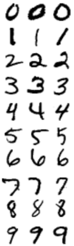
$\lambda = 10^3$

$\lambda = 10^4$

Images to classify



Training set



Classification

- Training set of size L for K classes:

$$\mathcal{S} = \left\{ (u_\ell, z_\ell) \in \mathbb{R}^N \times \{1, \ldots, K\} \;\middle|\; \ell \in \{1, \ldots, L\} \right\}$$

<u>examples</u>: $u_\ell =$  and $z_\ell = 2$

$u_\ell =$  and $z_\ell = 9$

Learning: multiclass SVM

- $\phi(u) : \mathbb{R}^N \to \mathbb{R}^M$: mapping from the input space onto an arbitrary feature space with $M > N$

  $\Rightarrow$ linearization

  examples: convolution networks [Mirowski et al., 2008]

  scattering coefficients [Brunat,Mallat,2013]

- The predictor relies on K different discriminating functions $D_k : \mathbb{R}^N \to \mathbb{R}$:

$$D_k(u) = \phi(u)^\top x^{(k)} + b^{(k)}$$

- The predictor selects the class that best matches an observation

$$d(u) = \arg \max_{1 \leq k \leq K} D_k(u)$$

Learning: multiclass SVM

Objective of the learning stage: estimate x to correctly predict the input-output pair $(u_\ell, z_\ell) \in \mathcal{S}$ for every $\ell \in \{1, \ldots, L\}$,

$$z_\ell = \arg\max_{1 \le k \le K} \varphi(u_\ell)^\top \mathrm{x}^{(k)}$$

$$\Leftrightarrow \quad \max_{k \neq z_\ell} \varphi(u_\ell)^\top (\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) < 0$$

[relax the strict ineqality with $\mu_\ell > 0$] $\Leftrightarrow \max_{k \neq z_\ell} \varphi(u_\ell)^\top (\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \le -\mu_\ell$

[deal with unfeasible constraints $\zeta^{(\ell)} \ge 0$] $\Leftrightarrow \max_{k \neq z_\ell} \varphi(u_\ell)^\top (\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \le \zeta^{(\ell)} - \mu_\ell$

$$\underset{(\mathrm{x},\xi) \in \mathbb{R}^{(M+1)K} \times \mathbb{R}^L}{\text{minimize}} \quad \sum_{k=1}^{K} \|x^{(k)}\|_2^2 + \lambda \sum_{\ell=1}^{L} \xi^{(\ell)} \quad \text{subj. to}$$

$$\begin{cases} (\forall \ell \in \{1, \ldots, L\}) & \max_{k \neq z_\ell} \varphi(u_\ell)^\top (\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \le \xi^{(\ell)} - \mu_\ell \\ (\forall \ell \in \{1, \ldots, L\}) & \xi^{(\ell)} \ge 0, \end{cases}$$

Image deconvolution with CNN

- Inverse problems : Tikhonov penalization

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \|Hx - z\|^2 + \lambda \|\Gamma x\|_2^2$$

$$\Leftrightarrow \widehat{x} = (H^* H + \lambda \Gamma^* \Gamma)^{-1} H z = G z.$$

- Reformulation into a convolutional network using the kernel separability theorem relying on the existence of the decomposition $G = USV^\top$:

$$\widehat{x} = \sum_j s_j U_{j,\bullet} (V_{j,\bullet}^\top z).$$

  where s_j denotes the j-th singular value, and $U_{j,\bullet}$ (resp. $V_{j,\bullet}$) denotes the j-th column of U (resp. V).
- 2D deconvolution can be reformulated as a weighted sum of separable 1D filters.
- $\widehat{x}$ can be well approximated by a small number of separable filters by dropping out kernel associated with very small s_j.

Image deconvolution with CNN

- Image Deconvolution Convolutional Neural Networks (DCNN) [Xu et al, 2014] :

$$\begin{aligned}\widehat{x} &= f(z)\\ &= W_3\sigma(W_2\sigma(W_1 z + b_1) + b_2.\end{aligned}$$

  - W_3 denotes weights playing the same role than S,
  - W_2 and W_1: separable kernels acting horizontally or vertically,
  - σ denotes a nonlinear function.

- Goal: estimate $(W_i)_{i=1,2,3}$ and $(b_i)_{i=1,2}$ in order to minimize

$$\frac{1}{2|N|}\sum_{i\in N}\|f(z_\ell) - \overline{x}_\ell\|.$$

using training image pairs $\{\overline{x}_\ell, z_\ell\}_{\ell\in N}$.