Scaled gradient projection methods for X-rays CT image reconstruction from reduced sampling

V.L. Coli[1], E. Loli Piccolomini[2], E. Morotti[3] and L. Zanni[1]

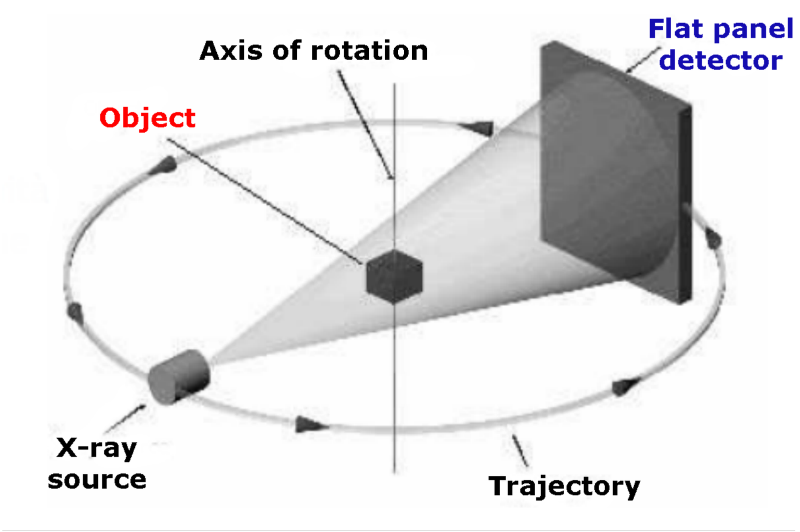
[1]University of Modena and Reggio Emilia , [2]University of Bologna, [3]University of Padova

3d tomography and problem formulation

Modern CT techniques are characterized by few 3d scans, performed by an X-ray source that wheels around the object of interest. Beyond this volume, a digital flat detector captures the X-ray cone beam, gathering 2d projected images. The 3d digital reconstruction of the volume is visually conceived as a stack of many 2d layers, overlaying in a prefixed direction.

The image formation process is modelled as an underdetermined linear system $Ax = b$ where:

- x is the unknown volume, discretized in $N := N_x \times N_y \times N_z$ voxels;
- b is the collection of all the noisy projections of x on the detector (made of $n_x \times n_y$ pixels), performed from n_θ angles taken in a semi-sphere ($n := n_x \times n_y \times n_\theta$ and $n \ll N$);
- A is a matrix of size $n \times N$, representing how the tomograph moves around x with respect to the detector.



The reconstructed image x is the solution of the constrained problem:

$$\min_{x\geq 0} \; f(x) = KL(x) + \lambda TV(x)$$

- $KL(x) = \sum_{j_x=1}^{N_x}\sum_{j_y=1}^{N_y}\sum_{j_z=1}^{N_z}[b\, log(\frac{b}{Ax+bg}) + (Ax+bg) - b]_{j_x,j_y,j_z}$ is the Kullback-Leibler data-fitting function, in case of Poisson-noise (and bg is the background value);
- $\lambda > 0$ is the regularization parameter;
- $TV(x) := \sum_{j=1}^{N}\left(\|D_j x\|_2^2 + \beta^2\right)^{\frac{1}{2}}$ is the differentiable discrete Total Variation of x (with $D_j x \in \mathbb{R}^3$ 3d-discrete gradient of voxel x_j).

Scaled Gradient Projection (SGP) algorithm

SGP algorithm

Initialize: $x_0 \geq 0$, $\delta, \sigma \in (0,1)$, $0 \leq \alpha_{min} \leq \alpha_0 \leq \alpha_{max}$, $D_0 \in \mathcal{D}_\rho$;
for $k = 0, 1, \ldots$
  $d_k = P_+(x_k - \alpha_k D_k \nabla f(x_k)) - x_k$; (scaled gradient projection step)
  $\lambda_k = 1$;
  while $f(x_k + \lambda_k d_k) > f(x_k) + \sigma\lambda_k \nabla f(x_k)^T d_k$
    $\lambda_k = \delta\lambda_k$; (backtracking step)
  end
  $x_{k+1} = x_k + \lambda_k d_k$;
  define the diagonal scaling matrix $D_{k+1} \in \mathcal{D}_\rho$; (scaling updating rule)
  define the step–length $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}]$; (step–length updating rule)
end

Scaling strategy*

The split gradient idea leads to the following updating rule:

$$D_k = diag\left(\min\left\{\rho, \max\left\{\frac{1}{\rho}, \frac{x_k}{V^{KL}(x_k) + \lambda V^{TV}(x_k)}\right\}\right\}\right), \quad \rho > 1$$

where

$$\nabla f(x) = \nabla KL(x) + \lambda\, \nabla TV(x) = [V^{KL}(x) - U^{KL}(x)] + \lambda[V^{TV}(x) - U^{TV}(x)]$$

with $V^{KL}, V^{TV} > 0$ and $U^{KL}, U^{TV} \geq 0$

When $D_k = I$, the SGP remains a Gradient Projection method (GP).

* [Lantéri-Roche-Aime, *Inverse Problems*, 2002]

Step-length rule based on alternating Barzilai-Borwein steps†

Denoting with $s_{k-1} = (x_k - x_{k-1})$ and $z_{k-1} = (\nabla f(x_k) - \nabla f(x_{k-1}))$, compute

$$\alpha_k^{BB1} = \frac{{s_{k-1}}^T D_k^{-1} D_k^{-1} s_{k-1}}{{s_{k-1}}^T D_k^{-1} z_{k-1}} \quad \text{and} \quad \alpha_k^{BB2} = \frac{{s_{k-1}}^T D_k z_{k-1}}{{z_{k-1}}^T D_k D_k z_{k-1}},$$

and choose

$$\alpha_k = \begin{cases} min\{\alpha_j^{BB2} \mid j = max\{1, k - m_{BB}\}, \ldots, k\} & \text{if } \frac{\alpha_k^{BB2}}{\alpha_k^{BB1}} < \tau \\ \alpha_k^{BB1}, & \text{otherwise} \end{cases}$$

Step-length rule based on Ritz-like values*

Choose the step-lengths for m_R next iterations as

$$\alpha_{k-1+i} = \frac{1}{\theta_i}, \quad i = 1, \ldots, m_R$$

where θ_i, $i = 1, \ldots, m_R$ are the eigenvalues of an $m_R \times m_R$ tridiagonal matrix T derived by the last m_R scaled gradient steps:

$$\left[D_{k-m_R}^{\frac{1}{2}}\tilde{g}^{(k-m_R)}, \ldots, D_{k-1}^{\frac{1}{2}}\tilde{g}^{(k-1)}\right], \quad \tilde{g}_i^{(k-j)} = \begin{cases} 0 & \text{if } (x_{k-j})_i = 0 \\ \nabla f(x_{k-j})_i & \text{if } (x_{k-j})_i > 0 \end{cases}$$

In case of gradient methods for $\min \frac{1}{2}x^T Ax - b^T x$, T is given by m_R steps of the Lanczos process applied to A and starting from $g^{(k-m_R)}/\|g^{(k-m_R)}\|$; θ_i are the so-called *Ritz values* (estimates of the eigenvalues of A).

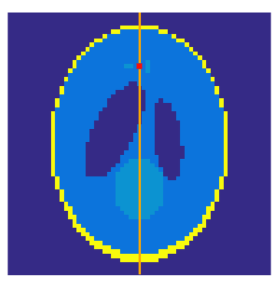
† [Zanella et al., *Inverse Problems*, 2009]
* [Fletcher, *Math. Program.*, 2012], [Porta-Prato-Zanni, *J. Sci. Comp.*, 2015]

Numerical results

Parameters:

- $N_x = N_y = N_z = 61$
- $n_x = n_y = 61$, $n_\theta = 37$
- SNR = 41.8257
- $\lambda = 0.03$, $\beta = 0.01$
- $m_R = m_{BB} = 3$
- $\delta = 0.4$, $\sigma = 2e-4$
- $\alpha_{min} = 1e-10$, $\alpha_{max} = 1e5$
- $\tau = 5e-1$

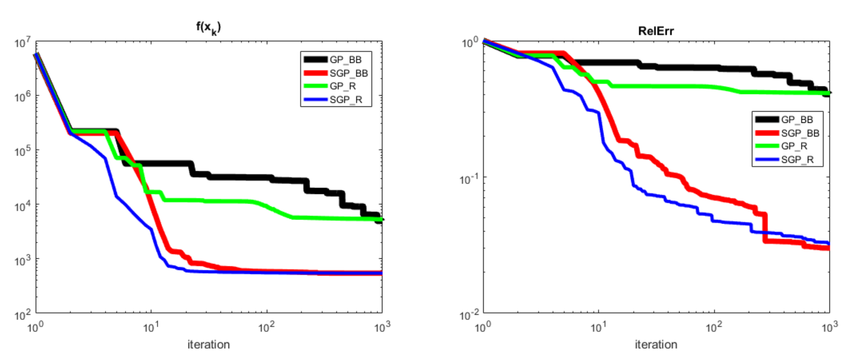
Test phantom:



Methods:

"GP_BB" is the non-scaled algorithm implemented with the Barzilai-Borwein rule, while "GP_R" is the non-scaled algorithm with Ritz rule. "SGP_BB" and "SGP_R" are the corresponding scaled versions.

Analysis:



	k=20		k=1000	
	$f(x_k)$	RelErr	$f(x_k)$	RelErr
SGP_BB	1055.09	0.1705	543.907	0.0301
SGP_R	590.133	0.0856	544.602	0.0324

Reconstructions and profiles of interest:

	SGP with BB rule	SGP with Ritz rule	Profiles along the vertical direction	Profiles along the depth direction
k=20				
k=1000