I. Experiments

A. Approximation of Toeplitz matrix by oversampled circulant matrix

For any (rectangular, symmetric) subset $S \subseteq \mathbb{Z}^n$, define $\mathbb{C}(S)$ to be the space of complex vectors $\mathbf{x}$ of length $|S|$ indexed by $\mathbf{k} \in S$. When it is clear from context, we will use $\mathbf{x}$ to denote an element of both $\mathbb{C}^{|S|}$ and $\mathbb{C}(S)$, and $\mathbf{x}[\mathbf{k}]$ for the entry of $\mathbf{k}$ at index $\mathbf{k} \in S$.

Suppose we want to recover a data vector $\mathbf{x} \in \mathbb{C}(\Delta)$ from samples $P_\Gamma \mathbf{x} = \mathbf{b} \in \mathbb{C}(\Gamma)$. And suppose we have an annihilating filter $\mathbf{d} \in \mathbb{C}(\Lambda)$.

Define $\mathcal{Q}\mathbf{x} = \mathcal{T}(\mathbf{x})\mathbf{d} = \mathcal{P}_{\Delta|\Lambda}[\mathbf{d} * \mathbf{x}] = \mathcal{P}_{\Delta|\Lambda}\mathcal{C}_\Delta \mathbf{x}$, where $\mathcal{C}_\Delta$ is circular convolution with $\mathbf{d}$ on a grid of size Δ, and $\mathcal{P}_S$ denotes projection onto S. In the CG step we minimize problems of the form

$$\min_{\mathbf{x}\in\mathbb{C}(\Delta)} \|\mathcal{Q}\mathbf{x}\|_2^2 \quad s.t. \quad \mathcal{P}_\Gamma \mathbf{x} = \mathbf{b}.$$

Writing $\mathbf{x} = \mathcal{P}_\Gamma^* \mathbf{b} + \mathcal{P}_{(\Delta-\Gamma)}^* \mathbf{y}$, it is equivalent to solve

$$\min_{\mathbf{y}\in\mathbb{C}(\Delta-\Gamma)} \|\mathcal{Q}\mathcal{P}_\Gamma^*\mathbf{b} + \mathcal{Q}\mathcal{P}_{(\Delta-\Gamma)}^*\mathbf{y}\|_2^2$$

which reduces to the linear system

$$\mathbf{y} \in \mathbb{C}(\Delta - \Gamma): \quad (\mathcal{P}_{(\Delta-\Gamma)}\mathcal{R}\mathcal{P}_{(\Delta-\Gamma)}^*)\mathbf{y} = -\mathcal{P}_{(\Delta-\Gamma)}^*\mathcal{R}\mathcal{P}_\Gamma^*\mathbf{b}.$$

Here $\mathcal{R} = \mathcal{Q}^*\mathcal{Q} = \mathcal{C}_\Delta^*\mathcal{P}_{\Delta|\Lambda}^*\mathcal{P}_{\Delta|\Lambda}\mathcal{C}_\Delta$. We want to circumvent the projection step $\mathcal{P}_{\Delta|\Lambda}^*\mathcal{P}_{\Delta|\Lambda}$ in order to speed up applications of $\mathcal{R}$, and to allow for the sum-of-squares simplification when there are multiple annihilating filters. Hence, we look at a circulant approximation $\mathcal{R}_{\Delta'} = \mathcal{C}_{\Delta'}^*\mathcal{C}_{\Delta'}$ where $\Delta' \subseteq \Delta$ is some "oversampled" grid where we perform the FFTs, and then project back down to Δ. Specifically, we solve

$$\mathbf{z} \in \mathbb{C}(\Delta' - \Gamma): \quad (\mathcal{P}_{(\Delta'-\Gamma)}\mathcal{R}_{\Delta'}\mathcal{P}_{(\Delta'-\Gamma)}^*)\mathbf{z} = -\mathcal{P}_{(\Delta'-\Gamma)}\mathcal{R}_{\Delta'}\mathcal{P}_\Gamma^*\mathbf{b}.$$

and then obtain our approximate solution $\widetilde{\mathbf{y}} \in \mathbb{C}(\Delta-\Gamma)$ as $\widetilde{\mathbf{y}} = \mathcal{P}_{(\Delta-\Gamma)}\mathbf{z}$ and we set $\widetilde{\mathbf{x}} = \mathcal{P}_\Gamma^*\mathbf{b} + \mathcal{P}_{(\Delta-\Gamma)}^*\widetilde{\mathbf{y}}$. The goal is to show

$$\|\mathbf{x} - \widetilde{\mathbf{x}}\| = \|\mathbf{y} - \widetilde{\mathbf{y}}\| \to 0$$

as the size of oversampling grid Δ' grows arbitrarily large. We also consider the (more realistic) problem where we relax the data constraint:

$$\min_{\mathbf{x}\in\mathbb{C}(\Delta)} \|\mathcal{Q}\mathbf{x}\|_2^2 + \lambda\|\mathcal{P}_\Gamma\mathbf{x} - \mathbf{b}\|_2^2.$$

which reduces to the linear system

$$\mathbf{x} \in \mathbb{C}(\Delta): \quad (\mathcal{C}_\Delta^*\mathcal{P}_{\Delta|\Lambda}^*\mathcal{P}_{\Delta|\Lambda}\mathcal{C}_\Delta + \lambda\mathcal{P}_\Gamma^*\mathcal{P}_\Gamma)\mathbf{x} = \lambda\mathcal{P}_\Gamma^*\mathbf{b}.$$

and its approximated version

$$\mathbf{z} \in \mathbb{C}(\Delta') : \quad (\mathcal{C}_{\Delta'}^* \mathcal{C}_{\Delta'} + \lambda \mathcal{P}_\Gamma^* \mathcal{P}_\Gamma)\mathbf{z} = \lambda \mathcal{P}_\Gamma^* \mathbf{b}.$$

with $\widetilde{\mathbf{x}} = P_\Delta(\mathbf{z})$.