Accelerated Alternating Optimization of MIMO Transceivers

Zilu Zhao, Fangqing.Xiao, Dirk Slock
Communication Systems Department, EURECOM, France
{zilu.zhao@eurecom.fr,fangqing.xiao,dirk.slock}@eurecom.fr

***Abstract*—Various majorization approaches allow to transform a Weighted Sum Rate criterion into a cost function that is quadratic in Tx or in Rx and is simple to optimize in terms of other parameters. The standard strategy is then to apply alternating minimization to the majorizer. These alternating approaches converge fairly slowly. Another issue that they typically are guaranteed to converge, but to a local optimum. Deterministic annealing has been proposed to find the global optimum. To reduce convergence time, acceleration techniques can be introduced. In this paper we review existing acceleration methods. Two big families arise, either Nesterov acceleration or Successive Over-Relaxation. The question arises and remains open which approach would be best suited to alternating precoder design.**

I. Introduction

A recent work on Transmitter (TX)/Receiver (Rx) design for multi-user Multi-Input Multi-Output (MIMO) systems can be found in [1].

II. Chebyshev and Polyak's Heavy Ball Methods

This reviews is largely based on [2].

A. Quadratic Optimization - Gradient Method

Quadratic cost function $f(\boldsymbol{x})$

$$\min_{\boldsymbol{x}}\{f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\,\boldsymbol{x} - \boldsymbol{x}^T\mathbf{b}\}\ ,\ \ \boldsymbol{x}_* = \boldsymbol{H}^{-1}\mathbf{b} \qquad (1)$$

Gradient Method (Steepest Descent)

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \gamma\,\nabla f(\boldsymbol{x}_{k-1}) = (\boldsymbol{I} - \gamma\,\boldsymbol{H})\,\boldsymbol{x}_{k-1} + \gamma\,\mathbf{b} \qquad (2)$$

Convergence

$$\boldsymbol{x}_k - \boldsymbol{x}_* = P_k^{grad}(\boldsymbol{H})\,(\boldsymbol{x}_0 - \boldsymbol{x}_*)\ ,\ \ P_k^{grad}(\boldsymbol{H}) = (\boldsymbol{I} - \gamma\,\boldsymbol{H})^k \qquad (3)$$

Minmax optimization: best worst-case convergence rate

$$\begin{aligned}\|\boldsymbol{I} - \gamma\,\boldsymbol{H}\| \ &\leq \max_{\mu\boldsymbol{I}\leq\boldsymbol{H}\leq L\boldsymbol{I}} \|\boldsymbol{I} - \gamma\,\boldsymbol{H}\| \\ &= \max_{\mu\leq\lambda\leq L} |1-\gamma\,\lambda| = \max\{\gamma\,L - 1, 1 - \mu\,\gamma\}\end{aligned} \qquad (4)$$

$$\min_{\gamma}\max\{\gamma\,L - 1, 1-\mu\,\gamma\} = \frac{L-\mu}{L+\mu}\ ,\ \gamma = \frac{2}{L+\mu} \qquad (5)$$

$$\|\boldsymbol{x}_k - \boldsymbol{x}_*\| \leq \left(\frac{\kappa-1}{\kappa+1}\right)^k \|\boldsymbol{x}_k - \boldsymbol{x}_*\| \qquad (6)$$

condition number $\kappa = \frac{L}{\mu} \geq 1$ of $f(.)$.

B. Chebyshev Method

First-order methods and monic matrix polynomials $P_k(0) = 1$:

$$\begin{aligned}&\boldsymbol{x}_{k+1} \in \boldsymbol{x}_0 + \text{span}\{\nabla f(\boldsymbol{x}_0), \nabla f(\boldsymbol{x}_1), \ldots, \nabla f(\boldsymbol{x}_k)\} \\ &\text{iff} \quad \boldsymbol{x}_k - \boldsymbol{x}_* = P_k(\boldsymbol{H})\,(\boldsymbol{x}_0 - \boldsymbol{x}_*)\,\forall k \geq 0\end{aligned} \qquad (7)$$

Conjugate gradient methods - Krylov subspaces. Minimax design over $\mathcal{M}\{\boldsymbol{H} : \mu\,\boldsymbol{I} \leq \boldsymbol{H} \leq L\,\boldsymbol{I}\}$

$$P_k^* = \arg\min_{P\in\mathcal{P}_k}\max_{\boldsymbol{H}\in\mathcal{M}} \|P(\boldsymbol{H})\| = \arg\min_{P\in\mathcal{P}_k}\max_{\lambda\in[\mu,L]} \|P(\lambda)\| \qquad (8)$$

Solution: (shifted) Chebyshev poynomials of the first kind: recursively

$$\begin{aligned}&\mathcal{T}_0(x) = 1\ ,\ \mathcal{T}_1(x) = x\ , \\ &\mathcal{T}_k(x) = 2x\,\mathcal{T}_{k-1}(x) - \mathcal{T}_{k-2}(x)\ ,\ \ k\geq 2\end{aligned} \qquad (9)$$

Shifting $[-1,1] \to [\mu, L]$, $\mathcal{T}_k(x) \to \mathcal{C}_k(x)$, we get

$$\boldsymbol{x}_k - \boldsymbol{x}_* = \mathcal{C}_k(\boldsymbol{H})\,(\boldsymbol{x}_0 - \boldsymbol{x}_*) \qquad (10)$$

Recall $\nabla f(\boldsymbol{x}) = \boldsymbol{H}\,(\boldsymbol{x} - \boldsymbol{x}_*)$.

C. Chebyshev and Polyak's Heavy Ball Methods

Using Chebyshev polynomial recursion:

$$\begin{aligned}&\delta_1 = \tfrac{\kappa-1}{\kappa+1}\ ,\ \boldsymbol{x}_1 = \boldsymbol{x}_0 - \tfrac{2}{L+\mu}\,\nabla f(\boldsymbol{x}_0) \\ &\text{for } k\geq 2: \\ &\quad \delta_k = (2/\delta_1 - \delta_{k-1})^{-1} \\ &\quad \boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \tfrac{4\delta_k}{L-\mu}\,\nabla f(\boldsymbol{x}_{k-1}) + (2\,\delta_k/\delta_1 - 1)(\boldsymbol{x}_{k-1} - \boldsymbol{x}_{k-2})\end{aligned} \qquad (11)$$

Last term = momentum term. Inherits 2^{nd}-order recursion from Chebyshev. Steady-state: $\delta_\infty = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$.

Polyak's Heavy Ball method: replace δ_k by δ_∞ in Chebyshev method:

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \frac{4}{(\sqrt{L}-\sqrt{\mu})^2}\,\nabla f(\boldsymbol{x}_{k-1}) + \delta_\infty^2\,(\boldsymbol{x}_{k-1} - \boldsymbol{x}_{k-2}) \qquad (12)$$

D. Worst Case Convergence Bounds Chebyshev

From $\boldsymbol{x}_k - \boldsymbol{x}_* = \mathcal{C}_k(\boldsymbol{H})\,(\boldsymbol{x}_0 - \boldsymbol{x}_*)$ we gt

$$\|\boldsymbol{x}_k - \boldsymbol{x}_*\| \leq \|\mathcal{C}_k(\boldsymbol{H})\,(\boldsymbol{x}_0 - \boldsymbol{x}_*)\| \leq \|\boldsymbol{x}_0 - \boldsymbol{x}_*\| \max_{\lambda\in[\mu,L]} |\mathcal{C}_k(\lambda)| \qquad (13)$$

$\max_{\lambda\in[\mu,L]} |\mathcal{C}_k(\lambda)| = |\mathcal{C}_k(L)|$. From which

$$\|\boldsymbol{x}_k - \boldsymbol{x}_*\| \leq \frac{2}{\xi + k + \xi^{-k}}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\| \leq 2\,\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \|\boldsymbol{x}_0 - \boldsymbol{x}_*\| \qquad (14)$$

where $\xi = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$. Last upperbound also shared by Heavy Ball method.

Compared to gradient method: $\kappa = L/\mu$ replaced by $\sqrt{\kappa}$.

Chebyshev requires $\sqrt{\kappa/2}$ fewer iterations to reach a certain convergence precision, compared to the gradient method.

III. Acceleration for L-Smooth

A. From Quadratic to L-Smooth, μ-Strongly Convex

$f(\boldsymbol{x})$ is L-smooth convex iff $\forall \boldsymbol{x}, \boldsymbol{y}$

$$f(\boldsymbol{x}) \leq f(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{y})^T \nabla f(\boldsymbol{y}) + \frac{L}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2 \tag{15}$$

$f(\boldsymbol{x})$ is μ-strongly convex iff $\forall \boldsymbol{x}, \boldsymbol{y}$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{y})^T \nabla f(\boldsymbol{y}) + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2 \tag{16}$$

Interpolation inequality for L-smooth, μ-strongly convex ($\mu \geq 0$):

$$\begin{aligned} f(\boldsymbol{x}) \geq\ & f(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{y})^T \nabla f(\boldsymbol{y}) + \tfrac{1}{2L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \\ & + \tfrac{\mu L}{2(L-\mu)}\|\boldsymbol{x}-\boldsymbol{y} - \tfrac{1}{L}(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))\|^2 \end{aligned} \tag{17}$$

including possibly $\mu = 0$. $\mathcal{F}_{\mu,L}$ family of L-smooth, μ-strongly convex f.

B. Lyapunov Function for Gradient Method

With $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \frac{1}{L}\nabla f(\boldsymbol{x}_{k-1})$ and $A_k = A_{k-1}+1$, for L-smooth f

$$\phi_k \leq \phi_{k-1} \ \text{ for } \ \phi_k = A_k(f(\boldsymbol{x}_k) - f_*) + \frac{L}{2}\|\boldsymbol{x}_k - \boldsymbol{x}_*\|^2 \tag{18}$$

ϕ_k is called a Lyapunov (or *potential* or *energy*) function. With $A_0 = 0$ we get $A_k = k$ and

$$f(\boldsymbol{x}_k) - f_* \leq \frac{L}{2k}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \tag{19}$$

Hence the gradient method produces a worst case convegence rate of $O(1/k)$ for $f \in \mathcal{F}_{0,L}$.

C. Optimized Gradient Method (OGM)

Consider

$$\boldsymbol{y}_k = \boldsymbol{y}_{k-1} + \sum_{i=0}^{k-1} h_{k,i}\nabla f(\boldsymbol{y}_i)\,, \text{ initialization } \boldsymbol{y}_0 \tag{20}$$

$$\min_{\{h_{j,i}\}} \max_{f \in \mathcal{F}_{0,L}} \frac{f(\boldsymbol{y}_k) - f_*}{\|\boldsymbol{y}_0 - \boldsymbol{x}_*\|^2} \tag{21}$$

Optimal Solution recursive: OGM [KimFessler'16] : $i \to k,\ k \to K$

$\boldsymbol{y}_0 = \mathbf{z}_0 = \boldsymbol{x}_0$, $\theta_{-1,K} = 0$

$$\begin{aligned} \theta_{k,K} &= \tfrac{1}{2} + \sqrt{\theta_{k-1,K}^2 + \tfrac{1}{4}} \\ \boldsymbol{y}_k &= (1 - \tfrac{1}{\theta_{k,K}})\,\boldsymbol{x}_k + \tfrac{1}{\theta_{k,K}}\,\mathbf{z}_k \\ \boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \tfrac{1}{L}\nabla f(\boldsymbol{y}_k) \\ \mathbf{z}_{k+1} &= \mathbf{z}_k - \tfrac{2\theta_{k,K}}{L}\nabla f(\boldsymbol{y}_k) \end{aligned} \tag{22}$$

with output approximation $\boldsymbol{y}_K = (1 - \frac{1}{\theta_{K,K}})\,\boldsymbol{x}_K + \frac{1}{\theta_{K,K}}\,\mathbf{z}_K$, $\theta_{K,K} = \frac{1}{2} + \sqrt{2\theta_{K-1,K}^2 + \frac{1}{4}}$. Equivalently (eliminate $\mathbf{z}_k$): $\boldsymbol{x}_{k+1} = \boldsymbol{y}_k - \frac{1}{L}\nabla f(\boldsymbol{y}_k)$ and

$$\boldsymbol{y}_{k+1} = \boldsymbol{x}_{k+1} + \frac{\theta_{k,K}-1}{\theta_{k+1,K}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \frac{\theta_{k,K}}{\theta_{k+1,K}}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k) \tag{23}$$

D. OGM Lyapunov Function and Convergence

Lyapunov function

$$\phi_k = 2\theta_{k-1,K}^2(f(\boldsymbol{y}_{k-1}) - f_* - \frac{1}{2L}\|\nabla f(\boldsymbol{y}_{k-1})\|^2) + \frac{L}{2}\|\mathbf{z}_k - \boldsymbol{x}_*\|^2 \tag{24}$$

Can show $\theta_{k,K} \geq \theta_{k-1,K} + \frac{1}{2} \geq \frac{k}{2} + 1$, $\theta_{K,K} \geq \frac{K+1}{\sqrt{2}}$. Worst case convergence speed for OGM with L-smooth f

$$\theta_{K,K}^2(f(\boldsymbol{y}_K) - f_*) \leq \phi_{K+1} \leq \phi_K \leq \cdots \leq \phi_0 = \frac{L}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \tag{25}$$

$$\text{Hence } \ f(\boldsymbol{y}_K) - f_* \leq \frac{L\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2}{(K+1)^2} \tag{26}$$

quadratic convergence speed compared to gradient: accelerated.

$\theta_{k,K}$ recursion comes from requiring canceling terms to get descent of $\phi_k \leq \phi_{k-1}$.

E. Nesterov Acceleration

Inspired by OGM, consider $\phi_k = A_k(f(\boldsymbol{x}_k) - f_*) + \frac{L}{2}\|\mathbf{z}_k - \boldsymbol{x}_*\|^2$ and generic accelerated gradient algorithm

$$\begin{aligned} \boldsymbol{y}_k &= \boldsymbol{x}_k + \tau_k(\mathbf{z}_k - \boldsymbol{x}_k) = \tau_k\,\mathbf{z}_k + (1-\tau_k)\,\boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \alpha_k \nabla f(\boldsymbol{y}_k) \\ \mathbf{z}_{k+1} &= \mathbf{z}_k - \gamma_k \nabla f(\boldsymbol{y}_k) \end{aligned} \tag{27}$$

Choosing $a_k = \frac{1}{2} + \sqrt{A_k + \frac{1}{4}}$, $A_{k+1} = A_k + a_k$, $\tau_k = a_k/A_{k+1}$, $\alpha_k = \frac{1}{L}$, $\gamma_k = a_k/L$, get Nesterov.

Can show $A_k \geq k^2/4$ for $A_0 = 0$. Convergence speed

$$\begin{aligned} & A_k(f(\boldsymbol{x}_k) - f_*) \leq \phi_k \leq \cdots \leq \phi_0 = \tfrac{L}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \\ \Rightarrow\ \ & f(\boldsymbol{x}_k) - f_* \leq \tfrac{L}{2A_k}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \leq \tfrac{2L}{k^2}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \end{aligned} \tag{28}$$

Factor 2 suboptimality compared to OGM. More usual equivalent form: gradient update + momentum

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \tfrac{1}{L}\nabla f(\boldsymbol{y}_k) \\ \boldsymbol{y}_{k+1} &= \boldsymbol{x}_{k+1} + \tfrac{a_k - 1}{a_{k+1}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \end{aligned} \tag{29}$$

IV. Acceleration for μ-Strongly Convex

A. Gradient Descent with μ-Strong Convexity

With $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \frac{1}{L}\nabla f(\boldsymbol{x}_{k-1})$ and now $A_k = (A_{k-1}+1)/(1-1/\kappa)$, consider now

$$\phi_k = A_k(f(\boldsymbol{x}_k) - f_*) + \frac{L + \mu A_k}{2}\|\boldsymbol{x}_k - \boldsymbol{x}_*\|^2 \tag{30}$$

A_k recursion has solution $A_k = \kappa((1-1/\kappa)^{-k} - 1)$. Exponential convergence now

$$f(\boldsymbol{x}_k) - f_* \leq \frac{\mu\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2}{2((1-1/\kappa)^{-k} - 1)} \tag{31}$$

B. Acceleration with μ-Strong Convexity

Accelerated gradient algorithm becomes (now introduces a non-zero δ_k)

$$\begin{aligned} \boldsymbol{y}_k &= \boldsymbol{x}_k + \tau_k\,(\mathbf{z}_k - \boldsymbol{x}_k) \\ \boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \alpha_k\,\nabla f(\boldsymbol{y}_k) \\ \mathbf{z}_{k+1} &= (1-\delta_k/\kappa)\,\mathbf{z}_k + \delta_k/\kappa\,\boldsymbol{y}_k - \gamma_k\,\nabla f(\boldsymbol{y}_k) \end{aligned} \tag{32}$$

For Nesterov, choose

$$\begin{aligned} & A_{k+1} = (A_k + \tfrac{1}{2} + \sqrt{A_k + A_k^2/\kappa + \tfrac{1}{4}})/(1-1/\kappa)\,,\ A_0 = 0 \\ & \tau_k = \tfrac{(A_{k+1}-A_k)(1+A_k/\kappa)}{A_{k+1}+2A_kA_{k+1}/\kappa - A_k^2/\kappa} \\ & \delta_k = \tfrac{A_{k+1}-A_k}{1+A_{k+1}/\kappa}\,,\ \alpha_k = \tfrac{1}{L}\,,\ \gamma_k = \tfrac{\delta_k}{L} \end{aligned} \tag{33}$$

Lyapunov function $\phi_k = A_k(f(\boldsymbol{x}_k) - f_*) + \frac{L+\mu\,A_k}{2}\,\|\mathbf{z}_k - \boldsymbol{x}_*\|^2$
Now $A_{k+1} \geq A_k/(1-1/\sqrt{\kappa})$ leading to

$$f(\boldsymbol{x}_k) - f_* \leq \min\{\frac{2}{k^2}, (1-\frac{1}{\sqrt{\kappa}})^k\}\,L\,\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|^2 \tag{34}$$

replacing again κ by $\sqrt{\kappa}$ in gradient descent.

C. μ-Strongly Convex Acceleration with Constant Momentum

We have $A_k \to \infty$, $A_{k+1}/A_k \to (1-1/\sqrt{\kappa})^{-1}$ and hence $\tau_k \to (\sqrt{\kappa}+1)^{-1}$, $\delta_k \to \sqrt{\kappa}$. Leads to Nesterov algorithm with constant momentum

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \tfrac{1}{L}\,\nabla f(\boldsymbol{y}_k) \\ \boldsymbol{y}_{k+1} &= \boldsymbol{x}_{k+1} + \tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\,(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \end{aligned} \tag{35}$$

Leads to the same exponential convergence for μ-strongly convex f.

V. Accelerated Gradient Techniques for Transceiver Optimization

In [3] the Weighted Sum Rate in multi-user MIMO communications is majorized by a quadratic Weighted Sum Mean Squared Error (WSMSE) criterion. This quadratic cost function is then further majorized by an L-smooth quadratic cost function, by throwing in a further auxiliary variable. This majorization operation is of the following form:

$$\begin{aligned} \boldsymbol{x}^T\,\boldsymbol{H}\,\boldsymbol{x} - 2\boldsymbol{x}^T\mathbf{b} = \min_{\mathbf{z}}\{& L\,\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{x}^T(L\,\boldsymbol{I} - \boldsymbol{H})\,\mathbf{z} + \\ & \mathbf{z}^T(L\,\boldsymbol{I} - \boldsymbol{H})\,\mathbf{z} - 2\boldsymbol{x}^T\mathbf{b}\} \\ \Rightarrow\ \mathbf{z} = \boldsymbol{x}\,,\ \boldsymbol{x} = \mathbf{z} - \tfrac{1}{L}&(\boldsymbol{H}\,\mathbf{z} - \mathbf{b}) \end{aligned} \tag{36}$$

The optimization of this L-smooth majorizer (called "non-homgeneous Upperbound" in [3]) leads to the interpretation of a typical gradient update, with furthermore the stepsize adjusted accoring to the L-smooth property. This majorization interpretation of a gradient update seems to be original and is quite remarkable. The motivation in [3] for the majorization by replacing the Hessian by a multiple of identity is totally focused on lowering the computational complexity by avoiding a matrix inverse per update, However, one should remark that in the context of alternating minimization of an overall non-convex cost function, the majorization may have some effect on avoiding local optima since this majorization is not identical to but nevertheless related to deterministic annealing [4] in which the white noise variance would be artificially (substantially) increased, so that optimal Tx/Rx solutions would correspond to matched filters which are then globally optimal. In the end, [3] then introduces the standard Nesterov accelerated gradient technique to accelerate the gradient style update of the proposed majorizer. The simulations show that the multiple of identity based majorization by itself speeds up the convergence already quite a bit, and then the Nesterov acceleration adds further convergence acceleration,

In [5], a low complexity linear precoding method for extremely large-scale MIMO systems is proposed. The authors consider Regularized Zero-Forcing (R-ZF) precoders. The proposed low complexity approach for solving the linear equations defining the R-ZF precoder is based on Successive Over-Relaxation (SOR) [6], [7]

$$\begin{aligned} & \boldsymbol{H}\boldsymbol{x} = \mathbf{b}\,,\ \boldsymbol{H} = \boldsymbol{D} + \mathbf{L} + \mathbf{L}^T \\ & (\boldsymbol{D} + \omega\mathbf{L})\,\boldsymbol{x} = \omega\mathbf{b} - (\omega\mathbf{L}^T + (\omega-1)\boldsymbol{D})\,\boldsymbol{x}, \\ & \quad (\boldsymbol{D} + \omega\mathbf{L})^{-1}(\omega\mathbf{L}^T + (\omega-1)\boldsymbol{D}) \end{aligned} \tag{37}$$

where $\omega > 1$ is the so-called relaxation factor. Note that $\omega = 1$ corresponds to the Gauss-Seidel iterations, which correspond exactly to alternating minimization of a quadratic cost functon with the linear system as normal equations. Note also that SOR is guaranteed to converge for $\omega \in (0, 2)$. [5] combines SOR with Chebyshev acceleration. Note that since the matrix of coefficients is symmetric, it might be preferable to consider Symmetric SOR (SSOR) [8]

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \alpha_k\,\boldsymbol{P}^{-1}(\boldsymbol{H}\,\boldsymbol{x}_k - \mathbf{b}) \\ \boldsymbol{P}(\omega) &= \tfrac{\omega}{2-\omega}(\tfrac{1}{\omega}\boldsymbol{D} + \mathbf{L})\,\boldsymbol{D}^{-1}\,(\tfrac{1}{\omega}\boldsymbol{D} + \mathbf{L})^T) \end{aligned} \tag{38}$$

. It is known that for both SOR and SSOR, when an optimized relaxation factor ω_* is used, we have an improved conditioning of the ((S)SOR) preconditioned system $\kappa(\boldsymbol{P}^{-1}(\omega_*)\,\boldsymbol{H}) = O(\sqrt{\kappa(\boldsymbol{H})})$ [8], [9]. Which exhibits similarity to gradient acceleration. But ω_* is expensive to find. In any case, if an optimized relaxation factor can be used, it does not appear necessary to combine SOR with accelerated gradient techniques. Also, this seems to indicate that both accelerated gradient techniques and SOR (relaxed alternating minimization) are two different approaches to achieve accelerated iterative solutions with similar theoretical convergence speeds. Nevertheless, one may wonder if one of these approaches would be better suited for the alternating minimization appearing in WSR optimization. Note that in [5] SOR is applied to find one MIMO Tx precoder with low complexity. However, in MIMO systems, we typically have to alternate also at a higher level, between the optimization of the various Tx and Rx involved. The Kaczmarz approach is another algorithm in which one iteratively forces subsets of the equations to be satisfied, In [10] the authors consider random subset selection and they optimize the probability distribution of this selection. Expressions are derived for the expected convergence rate, which can apparently be improved significantly by randomness and distribution optimization. It is not clear though if this can be made as competitive as some of the alternatives above, considering some of the results in the next reference.

[11] refers to work by Nesterov on accelerated Gauss-Seidel, i.e. Block Coordinate Descent (Alternating Minimization). The authors introduce randomizing the block partitioning and updating order and frequency.The analysis on expected convergence rates show great potential improvement. In some of the simulations, the randomization of the updates leads to faster convergence than Nesterov acceleration. The combination of both ingredients leads to the fastest convergence of course. Their approach also applies to Kaczmarz.

VI. Stepsize Optimization

Stepsize optimization is an important issue in stochastic gradient techniques as applied to e.g. adaptive filtering (the Least Mean Squares (LMS) algorithm) [12]. In [13], adaptive stepsizes are derived for accelerated stochastic gradient. The resulting expressions are impressively complex though.

For the case of alternating minimization, some work on stepsize adaptation appears in [14] and associated papers. Acceleration for alternating majorization minimization is considered there. Related work appears in [15].

Acceleration has been applied for some time in compressive sensing, esp. for LASSO in the form of the Fast Iterative Shrinkage and Thresholding Algorithm (FISTA), where standard Nesterov acceleration is used. In [16], optimized stepsizes are proposed for successice convex approximations (SCAs) (majorization) and momentum terms, and extending stepsize optimization from line search to 2D stepsize subspace optimization. This brings up connections with the Affine Projection Algorithm (APA) which is a multi-dimensional projection generalization of the Normalized LMS algorithm [17].

VII. Concluding Remarks

In much of the state of the art (SotA), the beamforming design is not really optimized (e.g. take a regularized zero-forcing beamformer). Producing an optimized beamformer for e.g. sum rate takes an addiional complexity level w.r.t. R-ZF.

Tx beamformer design is a *constrained optimization* problem. Most of the SotA on accelerated algorithms are for unconstrained problems. Tx powers and Lagrange multipliers could be optimized separately. Not clear how much each optimization affects the overall convergence speed.

The *SotA on accelerated gradient techniques* considers extremely *little knowledge on the Hessian* of the optimization problem and hence comes up with quite *conservative worst case optimizations*. There appears to be *significant room to explore step-size optimization*.

Assuming we want an optimal (e.g. max WSR) precoder design, which iterative technique is fastest for given computational complexity: e.g. (accelerated) gradient or (accelerated) Gauss-Seidel/successive over-relaxation? Perhaps the optimal could be to move from gradient initially to Gauss-Seidel at convergence?

Another question is whether *deterministic annealing* with an optimized updating schedule and technique would be a contender for a low complexity solution with *furthermore global optimality*?

A sobering thought however in the context of precoder design is that *much simpler algorithms exist that compute approximate solutions and may require less channel state information*, e.g. beamspace, or pathwise processing (covariance CSIT), or naive UL/DL duality, regularized or reduced-order ZF,....

VIII. Acknowledgements

EURECOM' research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, by the Franco-German project CellFree6G, by the French PEPR-5G project PERSEUS, and by a Huawei France funded Chair towards Future Wireless Networks.

References

[1] A. Tibhirt, D. T. Slock, and Y.-W. Yi, "Multi-Cell Multi-User MIMO Imperfect CSI Transceiver Design with Power Method Generalized Eigenvectors," in *IEEE Asilomar Conf. Signals, Systems and Computers*, 2022.

[2] A. dAspremont, D. Scieur, and A. Taylor, *Acceleration Methods*. Foundations & Trends in Optimization, 2021.

[3] K. Shen, Z. Zhao, Y. Chen, Z. Zhang, and H. V. Cheng, "Accelerating Quadratic Transform and WMMSE," *IEEE J. Selected Areas in Communications*, Nov. 2024.

[4] F. Negro, S. Prasad Shenoy, I. Ghauri, and D. Slock, "On the MIMO Interference Channel," in *Proc. IEEE Information Theory and Applications workshop (ITA)*, San Diego, CA, USA, Feb. 2010.

[5] S. Berra, A. Benchabane, S. Chakraborty, K. Maruta, R. Dinis, and M. Beko, "A Low Complexity Linear Precoding Method for Extremely Large-Scale MIMO Systems," *IEEE Open J. Vehicular Tech.*, Dec. 2024.

[6] "Successive Over-Relaxation," https://en.wikipedia.org/wiki/Successive_over-relaxation.

[7] D. M. Young, "A Historical Overview of Iterative Methods," *Computer Physics Communications*, 1989.

[8] J. Dongarra, "Symmetric Successive Over-Relaxation Preconditioning," https://www.netlib.org/linalg/html_templates/node58.html, 1995.

[9] V. B. O. Axelsson, *Finite Element Solution of Boundary Value Problems: Theory and Computation*. SIAM, 2001.

[10] Z. Wang, C. Pan, Y. Huang, S. Jin, and G. Caire, "Randomized Iterative Algorithms for Distributed Massive MIMO Detection," *IEEE Trans. Sig. Proc.*, 2025.

[11] S. Tu, S. Venkataraman, A. C. Wilson, A. Gittens, M. I. Jordan, and B. Recht, "Breaking Locality Accelerates Block Gauss-Seidel," *arxiv:1701.03863*, Jan. 2017.

[12] D. T. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms," *IEEE Trans.Sig. Proc.*, Sept. 1993.

[13] Y. Yuan, D. H. K. Tsang, and V. K. N. Lau, "Step Size Adaptation for Accelerated Stochastic Momentum Algorithm Using SDE Modeling and Lyapunov Drift Minimization," *IEEE Trans. Signal Processing*, 2025.

[14] N. Gillis, "Inertial and Extrapolated Block Majorization Minimization with Application to NMF," https://www.dropbox.com/scl/fi/c8h4rcrcgf7ao4k1uzamj/TITIANandBMMwithExtrapol.pdf, 2024, keynote at NOPTA.

[15] M. Callahan, T. Vu, and R. Raich, "On Momentum Acceleration for Randomized Coordinate Descent in Matrix Completion," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[16] L. Schynol, M. Hemsing, and M. Pesavento, "An Accelerated Successive Convex Approximation Scheme With Exact Step Sizes for L1-Regression," *IEEE Open Journal of Sig. Proc.*, 2025.

[17] D. T. Slock, "The Block Underdetermined Covariance (BUC) Fast Transversal Filter (FTF) Algorithm for Adaptive Filtering," in *IEEE Asilomar Conf. Signals, Systems and Computers*, 1992.