Matrix-free MoM and current optimization

Mats Gustafsson

Electrical and Information Technology, Lund University, Sweden
Extinction cross section: spheres with radius $a = 50 \text{ nm}$

Extinction cross section $\sigma_{\text{ext}} = \sigma_a + \sigma_s$ sum rule

$$\frac{1}{\pi^2} \int_0^\infty \sigma_{\text{ext}}(\lambda) \, d\lambda = \hat{e} \cdot \gamma_e \cdot \hat{e} = 4\pi a^3$$

Identical areas under the curves for all metals and inner structure (e.g., hollow silver spheres) [Gus10]. (matlab function)
Optimization problems

Trade-off between bandwidth and efficiency (left) and maximum Q-factor (right)

maximize $\mathbf{I}^\mathsf{H} \mathbf{R}_r \mathbf{I}$
subject to $\mathbf{I}^\mathsf{H} \mathbf{X} \mathbf{I} = 0$
$\mathbf{I}^\mathsf{H} \mathbf{X}_w \mathbf{I} = 2 \bar{P}_w$
$\mathbf{I}^\mathsf{H} \mathbf{R}_\Omega \mathbf{I} = 2 \bar{P}_\Omega$.

maximize $\mathbf{I}^\mathsf{H} \mathbf{X}_w \mathbf{I}$
subject to $\mathbf{I}^\mathsf{H} \mathbf{X} \mathbf{I} = 0$
$\mathbf{I}^\mathsf{H} (\mathbf{R}_r + \mathbf{R}_\Omega) \mathbf{I} \leq 1$.

Self-resonance is enforced by equal electric and magnetic stored energies. Optimizing over the current $\mathbf{I}$ ($N \times 1$-matrix) with given positive semidefinite $N \times N$ matrices $\mathbf{R}_r$ (radiated power), $\mathbf{X}_w$ (stored energy), $\mathbf{X}$ (reactance), $\mathbf{R}_\Omega$ (ohmic losses).

These QCQPs are not convex so we need to reformulate them in convex (or some other solvable) form.
Plasmonics and volume solvers

- Volume current densities for optimization.
- Consider a cube with side length $\ell = N\Delta$.
- $\sim N^3$ unknowns with discretization $\Delta$.
- $N^3 = 10^6$ with $N = 100$.

Here,

- How can we solve EM and optimization problems with $N = 10^5$ to $N = 10^6$ unknowns.
Maximum gain, effective area, and absorption cross section

- **Solving the optimization problem**

  maximize $|\mathbf{FI}|^2 = \mathbf{I}^H \mathbf{U} \mathbf{I}$

  subject to $\mathbf{I}^H \mathbf{X} \mathbf{I} = 0$

  $\mathbf{I}^H (\mathbf{R}_\Omega + \mathbf{R}_r) \mathbf{I} = 1$,

- **Discretization (mesh) to describe geometry and current density.**
  - sufficient with $N = 1000$ to $N = 4000$ unknowns for small cases.
  - More for smaller $R_s$.
  - $N = 10^6$ to $N = 10^7$ for large cases.

How can we solve this for these large number of unknowns?

Maximum effective area $A_{\text{eff}} = G^2 \lambda^2 / (4\pi)$ for a rectangle with side lengths $\ell_x \times \ell_x/2$ and surface resistivity $R_s = 10^{-4}\eta_0$. 
Maximum gain, effective area, and absorption cross section

- Rewrite to an eigenvalue problem

\[ G_{\text{max}} = \text{minimize}_\nu \max \text{eig}(U, \nu X + R_\Omega + R_r) \]
\[ = \text{minimize}_\nu \max \text{eig}(F(\nu X + R_\Omega + R_r)^{-1}F^H) \]

- Here, \( \text{eig}(F(\nu X + R_\Omega + R_r)^{-1}F^H) \) is a \( 2 \times 2 \)-eigenvalue problem.
- Need to solve the linear system \( (\nu X + R_\Omega + R_r)^{-1}F^H \).
- \( R_\Omega \) is a sparse matrix with \( \approx 3N \) unknowns
- \( R_r = S^H S \) with \( S \) being \( M \times N \), where \( M \approx (ka)^2 \) the number of spherical modes.
- \( X \) a full \( N \times N \)-matrix. (\( N = 10^4 \rightarrow 0.8 \text{ GB} \) and \( N = 10^6 \rightarrow 8 \text{ TB} \))

Necessary to use matrix free algorithms to solve the linear system, i.e., iterative use of \( Ax \) without explicit matrix representation of \( A \).
Matrix free algorithms and BoR

- FFT-based
- Based on translational symmetry (Toeplitz matrix)
- Expend the Toeplitz matrix to a circular matrix.
- Diagonalize with FFT.
- Solve linear systems iteratively using FFT.
- Could also be used directly to solve eigenvalue problems, e.g., power iterations.

Fast multipole method (FMM)

- Use symmetries to reduce the size of the problem.
- Fourier series for BoR geometries.
- Only $m = 1$ for radiation in the $\hat{z}$-direction.
- Volume BoR MoM is similar to surface MoM.
Matrix-free algorithms

The computational complexity can be prohibitive for larger problems, e.g., the rather coarse discretization of $100 \times 100$ unknowns corresponds to a matrix $A$ with $10^8$ elements. Cannot store the matrix explicitly but can evaluate $Ax$ and $A^Hb$ efficiently. Often efficient to utilize (translational) symmetries and FFT based algorithms to reduce the computational complexity.

Using the same spacing for the basis functions and the data points gives a (block) Toeplitz matrix.
Toeplitz matrix vector multiplication

Let $A$ be an $M \times N$ Toeplitz matrix. Embed the Toeplitz matrix into a $(M + N) \times (M + N)$ circulant matrix such that

\[
\begin{pmatrix}
Ax \\
Sx
\end{pmatrix} = \begin{pmatrix}
A & S \\
S & A
\end{pmatrix}
\begin{pmatrix}
x \\
0
\end{pmatrix}
\]

The circulant matrix $\tilde{A}$ has the first row

\[
\tilde{A}_{1,:} = (A_{11} \ A_{12} \ \cdots \ A_{1N} \ A_{M1} \ \cdots \ A_{21})
\]

Evaluate using the FFT, i.e., from the first $M$ elements of, i.e.,

\[
Ax = [\mathcal{F}^{-1}(\mathcal{F}(\tilde{A}_{1,:})\mathcal{F}([x \ 0])))]_{1:M}
\]

Can reduce the dimension to $M + N - 1$

\[
\tilde{A}_{1,:} = (A_{11} \ A_{12} \ \cdots \ A_{1N} \ 0 \ A_{M1} \ \cdots \ A_{21})
\]

or $M + N + P - 1$ where $P \geq 0$ is the number of additional zeros.
Computational cost

The Pareto fronts are determined by solving

$$\text{maximize}_\nu \min \text{eig}(X_{\alpha\nu}, R_r),$$

which are concave in $\nu$ and $0 < \alpha < 1$ determines the Pareto front from

$$X_{\alpha\nu} = \alpha \nu X_e + \alpha (1 - \nu) X_m + (1 - \alpha) R_\Omega$$

- Need to sample $\alpha$ in 100-200 points for a smooth curve.
- 10-30 evaluations for $\text{max}_\nu$ (bisection algorithm).

In total $10^4$ to $10^5$ evaluations of generalized eigenvalue problems (similar to characteristic modes)

$$\min \text{eig}(X_{\alpha\nu}, R_r)$$
The real-valued part of the impedance matrix with elements

\[ Z_{pq} = jk Z_0 \int_\Omega \int_\Omega \psi_p(r_1) \cdot G(r_1, r_2) \cdot \psi_q(r_2) \, dS_1 \, dS_2 \]

is decomposed by expanding the Green dyadic in regular \( u^{(1)}_\alpha \) and out-going \( u^{(4)}_\alpha \) spherical vector waves

\[ G(r_1, r_2) = -jk \sum_\alpha u^{(1)}_\alpha(kr_<)u^{(4)}_\alpha(kr_>) , \]

where \( \alpha(\tau, \sigma, m, l) \), \( r_< = r_1 \) and \( r_> = r_2 \) if \( |r_1| < |r_2| \) and so on. Factorization

\[ R_r = S^T S = S^H S, \]

where \( S \) has the elements

\[ S_{\alpha p} = k Z_0^{1/2} \int_\Omega \psi_p(r) \cdot u^{(1)}_\alpha(kr) \, dS. \]

See [Tay+17] and compare with the T-matrix [Kri16], FMM [CRW93], and far-field expansion [GN13] methods.
Properties of $R = S^H S$

- $N_\psi$ and $N_\alpha \ll N_\psi$ number of basis function and spherical modes, respectively.
- $N_\alpha = 2L(L+2)$ with $L \approx \lceil ka + 3 + 7(ka)^{1/3} \rceil$, exponential convergence.
- Single (surface) integral (negligible computational cost).
- $R_r = S^H S \succeq 0$ (in theory and practice).
- Radiated field expanded in spherical modes $F = SI$.
- Radiated power $P_r = \frac{1}{2} \Gamma^H R \Gamma = \frac{1}{2} |SI|^2 = \frac{1}{2} \sum_\alpha |F_\alpha|^2$. 
Computational efficiency

Factorize radiation matrix $R_t = S^H S$ to rewrite the eigenvalue problem $\text{eig}(X_{\alpha \nu}, R_t)$

$$X_{\alpha \nu} I = \lambda R_t I = \lambda S^H S I$$

as

$$I = \lambda X_{\alpha \nu}^{-1} S^H S I \Rightarrow S I = \lambda S X_{\alpha \nu}^{-1} S^H S I$$

Implying an $N_{\alpha} \times N_{\alpha}$ eigenvalue problem

$$\lambda^{-1} = \text{eig}(S X_{\alpha \nu}^{-1} S^H)$$
Conclusions and status

- Can compute many/most cases for surface currents
  - Direct solvers for small problems (Q-factor).
  - Factorization $R_r = S^H S$ for computational efficiency.
  - Iterative solvers for large cases (gain, absorption)
- BoR volume MoM code (initial verification).
- FFT volume MoM (hopefully in the future).
References


