Lecture 4: DNN's parameters optimization

Plan of presentation

- Smalest Attractive Ellipsoid
- Parameter optimization as Optimization Problem under Bilinear Matrix inequalities (BMI)
- Transformation of BMI to Linear Matrix Inequalities (LMI)
- Scheme of numerical realization and MATLAB procedures

Smalest Attractive Ellipsoid

We will consider the optimization problem corresponding to the minimization of the "size" of the ellipsoid $E_0(P_{attr})$. When we speak about the "size" of an ellipsoid with a matrix P_{attr} we do not mean its volume. A volume of an ellipsoid (or, equivalently, its determinant) in fact is a bad function for the characterization of its "size" by two following reasons:
since

$$\boxed{\det P_{attr}^{-1} = \prod_{i=1}^{N} \lambda_i(P_{attr}^{-1}) \text{ and } r_i(P_{attr}) = \frac{1}{\sqrt{\lambda_i(P_{attr})}} = \sqrt{\lambda_i(P_{attr}^{-1})},}$$

where $\lambda_i(P_{attr}^{-1})$ $(i = 1, ..., N)$ are the eigenvalues of the inverse ellipsoid matrix P_{attr}^{-1} and $r_i(P_{attr})$ are the longitude of i-th semi-axises of the ellipsoid $E_0(P_{attr})$.

Volume as the detrminant - not a good characteristic

Fact

In view of this, we may conclude that minimization of $\det(P_{attr}^{-1})$ *is equivalent to minimization of its volume:*

$$\boxed{\mathrm{vol}(P_{attr}) = \det P_{attr}^{-1} = \prod_{i=1}^{n} r_i^2(P_{attr}).}$$

But, the product $\prod_{i=1}^{N} r_i^2(P_{attr})$ *admits to have a very large value of one of semi-axises, for example,* $r_{i_0}(P)$ *and all others may be very-very small! This exactly means that vol*(P_{attr}) *is a* **very bad quality characteristic**.

Trace of inverse matrices as a good characteristic

That's why the criterion $\mathrm{tr}(P_{attr}^{-1})$ is preferable since

$$\boxed{\mathrm{tr}\left\{P_{attr}^{-1}\right\} = \sum_{i=1}^{N} \lambda_i(P_{attr}^{-1}) \geq \max_{i=1,\ldots,N} \lambda_i(P_{attr}^{-1}) = \lambda_{\max}(P_{attr}^{-1}),}$$

and the minimization of $\mathrm{tr}\left\{P_{attr}^{-1}\right\}$ guarantees, at least, the minimization of its maximum eigenvalue, and hence, this guarantees the minimization of the corresponding maximal semi-axis

$$r_{\max}(P_{attr}^{-1}) = \sqrt{\lambda_{\max}(P_{attr}^{-1})}$$

of the given ellipsoid $\mathcal{E}_0(P_{attr})$.

Remark

Important to note from the numerical-computation point of view, that $\mathrm{tr}\left\{P_{attr}^{-1}\right\}$ *is a linear function of the matrix* P_{attr}^{-1} *and* $\det(P_{attr}^{-1})$ *is not!*

Optimization of DNNO parameters

Let us associate the optimal parameters of DNNO with the solution of the following matrix optimization problem

$$\boxed{\begin{array}{c} \operatorname{tr}\left\{P_{attr}^{-1}\right\} \to \inf\limits_{P>0,\, A,\, L,\, W_0^*,\, W_1^*,\, \alpha>0,\, \varepsilon>0} \\ \text{subject to the matrix constraint (2)} \\ S_{\alpha,\varepsilon}\left(P, A, L, W_0^*, W_1^*\right) < 0, \\ P > 0, \alpha > 0, \varepsilon > 0, \end{array}} \tag{1}$$

where

$$\boxed{S_{\alpha,\varepsilon}=\left[\begin{array}{cccc} \begin{array}{c} P\left(\frac{\alpha}{2}I_{n\times n}+A-LC\right)+ \\ \left(\frac{\alpha}{2}I_{n\times n}+A-LC\right)^{\intercal}P \end{array} & PL & PW_0^* & PW_1^* \\ L^{\intercal}P & \text{-}\varepsilon I_{m\times m} & 0 & 0 \\ \left(W_0^*\right)^{\intercal}P & 0 & \text{-}\varepsilon I_{k_\varphi\times k_\varphi} & 0 \\ \left(W_1^*\right)^{\intercal}P & 0 & 0 & \text{-}\varepsilon I_{k_\psi\times k_\psi} \end{array}\right] < 0} \tag{2}$$

Problem formulation in new variables

Let us introduce new matrix variables

$$\boxed{X := P > 0,\ Y := PA,\ Z := PL,\ Z_0 := PW_0^*,\ Z_1 := PW_1^*} \tag{3}$$

Then the optimization problem (1) can be rewritten as

$$\boxed{\operatorname{tr}\left\{\frac{\varepsilon\beta}{\alpha}X^{-1}\right\} \to \inf_{X>0,\, Y,\, Z,\, Z_0,\, Z_1,\, \alpha>0,\, \varepsilon>0}} \tag{4}$$

under the matrix constraint

$$\boxed{S_{\alpha,\varepsilon} = \begin{bmatrix} \begin{array}{c}\alpha X + Y + Y^{\intercal} \\ -ZC - C^{\intercal}Z^{\intercal}\end{array} & Z & Z_0 & Z_1 \\ Z^{\intercal} & -\varepsilon I_{m\times m} & 0 & 0 \\ Z_0^{\intercal} & 0 & -\varepsilon I_{k_\varphi \times k_\varphi} & 0 \\ Z_1^{\intercal} & 0 & 0 & -\varepsilon I_{k_\psi \times k_\psi} \end{bmatrix} < 0} \tag{5}$$

Important remark

Remark

Notice that the function

$$\operatorname{tr}\left\{P_{attr}^{-1}\right\} = \operatorname{tr}\left\{\frac{\varepsilon\beta}{\alpha}X^{-1}\right\}$$

is a function of X^{-1}*, but the matrix constrain* $S_{\alpha,\varepsilon}(X, Y, Z, Z_0, Z_1) < 0$ *(5) is the function of* X*. Some modifications of the problem are required.*

Schur compliment implementation

Theorem (Schur's complement)

Let S be a square matrix partitioned as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^{\intercal} & S_{22} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

where $S_{11} \in \mathbb{R}^{n\times n}$ is a symmetric $n \times n$ matrix and $S_{22} \in \mathbb{R}^{m\times m}$ is a symmetric $m \times m$ matrix. Then $S > 0$ **if and only if**

$$\boxed{\begin{array}{c} S_{11} > 0, \\ S_{22} > 0, \\ S_{11} - S_{12}S_{22}^{-1}S_{12}^{\intercal} > 0, \\ S_{22} - S_{12}^{\intercal}S_{11}^{-1}S_{12} > 0. \end{array} \right\}} \quad (6)$$

Nonnegative definiteness of a partitioned matrix

Theorem (Extended Schur's complement)

Let S be a square matrix partitioned as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^{\intercal} & S_{22} \end{bmatrix},$$

where $S_{11} = S_{11}^{\intercal} \in R^{n \times n}$, $S_{22} = S_{22}^{\intercal} \in R^{m \times m}$. *Then* $S \geq 0$ *if and only if*

$$\boxed{\begin{array}{c} S_{11} \geq 0, \ S_{22} \geq 0, \\ S_{11} S_{11}^{+} S_{12} = S_{12}, \ S_{22} S_{22}^{+} S_{12}^{\intercal} = S_{12}^{\intercal}, \\ S_{22} - S_{12}^{\intercal} S_{11}^{+} S_{12} \geq 0, \ S_{11} - S_{12} S_{22}^{+} S_{12}^{\intercal} \geq 0. \end{array}} \quad (7)$$

Here the H^{+} *is the matrix, pseudoinversed (in the Moore-Penrouse sence) to* H*, satisfying the identities*

$$HH^{+}H = H, H^{+}HH^{+} = H^{+}, \left(HH^{+}\right)^{\intercal} = HH^{+}, H^{+} = H^{\intercal}\left(HH^{\intercal}\right)^{+}.$$

Upper estimation of the minimization function

Using the upper estimate

$$\boxed{\begin{array}{c} X^{-1} \leq Q \Leftrightarrow 0 \leq Q - X^{-1} = 0 \leq Q - I_{n\times n} X^{-1} I_{n\times n} \\ S_{11}{=}Q,\ \ S_{22}^{+}{=} \left(X^{-1}\right)^{+} = \left(X^{-1}\right)^{-1} {=}X > 0,\ \ S_{12}{=}S_{12}^{\intercal}{=}I_{n\times n}^{-1} \end{array}} \tag{8}$$

for some matrix $Q > 0$, and in view of the Schir's complement we are able to represent the constraint (8) as

$$\boxed{\begin{bmatrix} Q & I_{n\times n} \\ I_{n\times n} & X \end{bmatrix} \geq 0} \tag{9}$$

Now let us take into account that the soltion of the problem (10) guarantees the solution of initial problem (4):

$$\boxed{\mathrm{tr}\left\{\frac{\varepsilon\beta}{\alpha}X^{-1}\right\} \leq \mathrm{tr}\left\{\frac{\varepsilon\beta}{\alpha}Q\right\} = \beta\mathrm{tr}\left\{\frac{\varepsilon}{\alpha}Q\right\} \to \inf_{Q>0,\ X>0,\ Y,\ Z,\ Z_0,\ Z_1,\ \alpha>0,\ \varepsilon>0}} \tag{10}$$

Optimization problem in new variables

$$\boxed{\operatorname{tr}\left\{\frac{\varepsilon\beta}{\alpha}Q\right\} \to \inf_{Q>0,\ X>0,\ Y,\ Z,\ Z_0,\ Z_1,\ \alpha>0,\ \varepsilon>0}} \tag{11}$$

$$\boxed{\begin{array}{c} S_{\alpha,\varepsilon} = \begin{bmatrix} \begin{array}{c}\alpha X + Y + Y^{\intercal} \\ -ZC - C^{\intercal}Z^{\intercal}\end{array} & Z & Z_0 & Z_1 \\ Z^{\intercal} & -\varepsilon I_{m\times m} & 0 & 0 \\ Z_0^{\intercal} & 0 & -\varepsilon I_{k_\varphi \times k_\varphi} & 0 \\ Z_1^{\intercal} & 0 & 0 & -\varepsilon I_{k_\psi \times k_\psi} \end{bmatrix} < 0 \\ \\ X > 0,\ \ Q > 0,\ \begin{bmatrix} Q & I_{n\times n} \\ I_{n\times n} & X \end{bmatrix} \geq 0 \end{array}} \tag{12}$$

Under fixed scalar parameters $\alpha > 0$, $\varepsilon > 0$ this is the matrix optimization problem with LMI constraints.

Algorithm:

1) At each step k $(k = 1, 2, \ldots)$ of iterations for any fixed positive scalars the constraints $\alpha_k > 0, \varepsilon_k > 0$ the matrix inequalities (12) becomes LMI's and the corresponding optimization problem can be effectively solved using appropriate mathematical software such as MATLAB with any SDP (Special Delivered Package) solver like **SEDUMI** or **YALMIP**. Let us denote by

$$g(\alpha_k, \varepsilon_k) := \min_{Q>0,\ X>0,\ Y,\ Z,\ Z_0,\ Z_1} \operatorname{tr}(Q), \begin{bmatrix} Q & I_{n\times n} \\ I_{n\times n} & X \end{bmatrix} > 0$$

the corresponding minimal value.

2) The optimization of the function $g(\alpha_k, \varepsilon_k)$ with respect to parameter α_k, ε_k can be realized locally basing on some derivative-free method, for example, using the MATLAB function fminsearch. In particular,

$$\alpha_{k+1} = \alpha_k + \Delta\alpha_k,\ \Delta\alpha_k > 0,\quad \varepsilon_{k+1} := \varepsilon_k - \Delta\varepsilon_k,\ \Delta\varepsilon_k > 0.$$

If ε_{k+1} becomes to be negative, we return back to the previous positive value. The same should be done if the matrix optimization problem says that admissible solutions do not exist.

3) Then iterations repeat.

Recuperation of the original matrices

Recall that

$$X := P > 0, Y := PA, Z := PL, Z_0 := PW_0^*, Z_1 := PW_1^*$$

So, if $Q^* > 0$, $X^* > 0$, Y^*, Z^*, Z_0^*, Z_1^*, $\alpha^* > 0$, $\varepsilon^* > 0$ are the solutions of the matrix optimization problem(11)-(12), then the original optimal matrices P^*, A^*, L^*, W_0^{**}, W_1^{**} may be found as

$$\boxed{\begin{array}{c} P^* = X^*,\ A^* = \left(X^*\right)^{-1} Y^*,\ L^* = \left(X^*\right)^{-1} Z^*, \\ \\ W_0^{**} = \left(X^*\right)^{-1} Z_0^*,\ W_1^{**} = \left(X^*\right)^{-1} Z_1^* \end{array}} \tag{13}$$

and

$$\boxed{P_{attr}^* = \frac{\alpha^*}{\varepsilon^* \beta} P^* = \frac{\alpha^*}{\varepsilon^* \beta} X^*} \tag{14}$$