# Online learning of the transmission matrix of dynamic scattering media

Lorenzo Valzania, Sylvain Gigan

# Online learning of the transmission matrix of dynamic scattering media

LORENZO VALZANIA* AND SYLVAIN GIGAN

*Laboratoire Kastler Brossel, École Normale Supérieure - Paris Sciences et Lettres (PSL) Research University, Sorbonne Université, Centre National de la Recherche Scientifique (CNRS) UMR 8552, Collège de France, 24 rue Lhomond, 75005 Paris, France*

*Corresponding author: lorenzo.valzania1@gmail.com

**Abstract:** Thanks to the latest advancements in wavefront shaping, optical methods have proven crucial to achieve imaging and control light in multiply scattering media, like biological tissues. However, the stability times of living biological specimens often prevent such methods from gaining insights into relevant functioning mechanisms in cellular and organ systems. Here we present a recursive and online optimization routine, borrowed from time series analysis, to optimally track the transmission matrix of dynamic scattering media over arbitrarily long timescales. While preserving the advantages of both optimization-based routines and transmission-matrix measurements, it operates in a memory-efficient manner. Because it can be readily implemented in existing wavefront shaping setups, featuring amplitude and/or phase modulation and phase-resolved or intensity-only acquisition, it paves the way for efficient optical investigations of living biological specimens.



## 1. Introduction

Optical methods are an irreplaceable tool to investigate biological media. They deliver images at numerous contrast mechanisms [1], and can activate injected biomolecules [2] and fluorescent markers [3]. However, precisely delivering light in space and time through biological tissues is not straightforward, as photons get multiply scattered by heterogeneities of tissues, limiting their penetration depth [4].

Another current challenge lies in tracking the scattering behaviour of living specimens, with decorrelation times up to only a few ms [5]. This proves crucial to understand the functioning mechanisms of cells and organisms, which requires their observation at extremely different timescales, from nanoseconds (at a molecular level) to minutes (for organ systems) [6]. The need for fast data acquisitions results, in turn, in measurements with inherently low signal-to-noise ratios (SNRs), and requires solving long and multidimensional time series [7], whose prohibitive size can make their evaluation problematic.

Wavefront shaping techniques have established themselves as the tools of choice to guide light in scattering media [8]. The transmission of arbitrary fields [9], point-spread-function (PSF) engineering [10], imaging [11], as well as tuning energy transmission through scattering media [12], become all accessible if the transmission or reflection matrix of the medium is measured [8,13]. In what follows, we will generically refer to transmission and reflection matrices as 'transfer matrices'. Conventional methods to retrieve the transfer matrix yield sub-optimal solutions in noisy environments [8]. Those optimization routines which can compensate for noise in the transfer matrix [14] require storing in memory the whole history of past measurements, making them unsuited with long streams of data.

Iterative, optimization-based, sequential algorithms to focus through scattering media yield an increase in the focus intensity already at their early iterations, which makes them the preferred option on dynamic media. Importantly, they are cast as *recursive* procedures, *i.e.*, computing the new estimate of the solution only requires the previous estimate and the new data point.

Unfortunately, their stochastic nature makes optimization over a set of output modes less reliable and the transmission of arbitrary fields prohibitive. Moreover, these procedures rely on maximizing a given metric, limiting light control to one predefined task. Various implementations derived from genetic algorithms [15, 16] have shown better resilience to noise than sequential algorithms, however at the cost of a higher computational complexity and careful choice of several adjustable parameters.

In signal processing, communications and finance, where most datasets are multidimensional time series, the recursive least-squares (RLS) algorithm has played a central role for system identification and prediction [17–19]. It allows optimal learning of linear predictors in an online manner—predictors are updated every time a new piece of data is sequentially made available, however past data do not need to be stored in memory. Consequently, its computational complexity is independent of the length of the time series, so iterations can be run over and over, ideally at the same rate as data acquisition (real-time operation).

Here, we demonstrate that the RLS algorithm represents a valuable tool to optimally estimate the transfer matrix of dynamic scattering media online and recursively. The least-squares optimization ensures resilience to noise. The algorithm is provided with a tunable memory, such that the dynamics of the scattering medium is accounted for. By doing so only the most reliable data points, *i.e.*, those acquired within the stability time of the medium, are used during the optimization. We justify how the RLS model can fit a wide variety of dynamic mechanisms happening in scattering media. Its performance is showcased with both simulated and experimental results, tracking the transmission matrix and the time-gated reflection matrix at realistic noise levels and well-controlled stability times, upon translating the scattering medium across the incident beam. We further show how light optimization can be achieved with binary amplitude or phase modulation and with phase-resolved or intensity-only measurements. Based on its computational complexity, we discuss its feasibility for light control in living biological specimens at large fields of view. Its simple implementation and the low number of adjustable parameters (whose choice is motivated in the next sections) make our proposed method readily applicable in existing wavefront shaping setups.

## 2. Methods

The method bears similarities with conventional routines for the measurement of the transfer matrix, and its working principle is graphically summarized in Fig. 1(a). However, here we allow the transfer matrix $\boldsymbol{X}_t \in \mathbb{C}^{M\times N}$ of the scattering medium to be dynamic, where we have denoted the number of output and input degrees of freedom with $M$ and $N$, respectively. At every time step $t$, while probing the medium with the input $\boldsymbol{a}_t \in \mathbb{C}^N$ and collecting the corresponding output $\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{a}_t \in \mathbb{C}^M$, we aim to solve the optimization problem $\hat{\boldsymbol{X}}_t = \arg\min_{\boldsymbol{X}_t} \mathcal{L}_t(\boldsymbol{X}_t)$, with

$$\mathcal{L}_t(\boldsymbol{X}_t) \equiv \sum_{\tau=1}^{t} \left(\lambda^{t-\tau}||\boldsymbol{y}_\tau - \boldsymbol{X}_t\boldsymbol{a}_\tau||^2\right) + \delta\lambda^t||\boldsymbol{X}_t||_F^2\,, \tag{1}$$

and where $||\cdot||$ and $||\cdot||_F$ denote the $L^2$-norm of a vector and the Frobenius norm of a matrix, respectively. Although for sake of generality the inputs and the outputs are assumed to be complex, we will also report an implementation where they are real, meaning that only the amplitude of the input beam is modulated and the intensity of the output fields is measured. Equation (1) is a linear least-squares loss function, featuring Tikhonov regularization via the regularization constant $\delta$. Note, however, that each *data-fidelity term* $||\boldsymbol{y}_\tau - \boldsymbol{X}_t\boldsymbol{a}_\tau||^2$ is exponentially weighted in time, such that the old pieces of data (corresponding to $\tau \ll t$) are less relevant than the most recent ones in the current estimation of the transfer matrix at time $t$. In other words, the *forgetting factor* $\lambda \leq 1$ endows the algorithm with a memory, which allows it to cope with dynamic transfer matrices—at every time step $t$, the optimization problem is solved anew, using

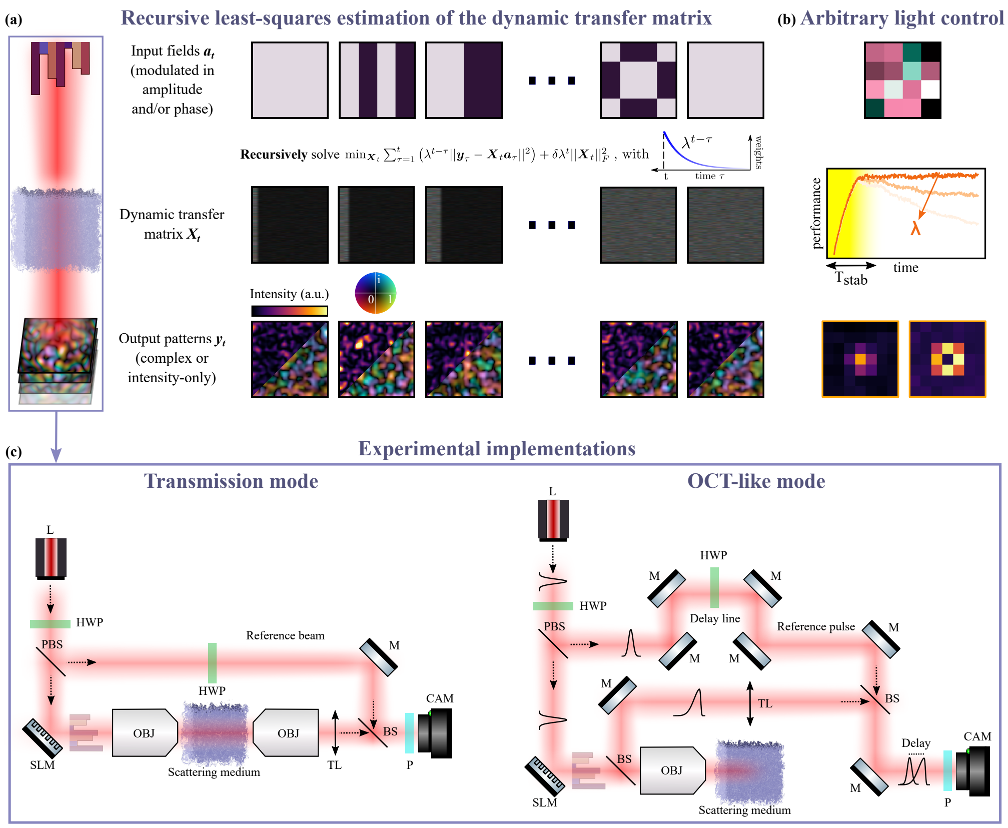


Fig. 1. Graphical summary of the RLS estimation technique and experimental implementations. (a) A sequence of input fields, modulated in amplitude and/or in phase (here Hadamard modulation patterns are shown), interacts with a dynamic scattering medium with unknown transfer matrix. Each (input, output) pair is used to update recursively the estimation of the dynamic transfer matrix, minimizing a regularized linear least-squares loss function, where each data-fidelity term is weighted via the coefficient $\lambda^{t-\tau}$. (b) At every time step, upon optimizing the coefficient $\lambda$, the current estimate of the transfer matrix can be used to achieve arbitrary light control through the scattering medium (here, a focus and a donut-shaped beam are displayed). (c) We demonstrated our method with a setup in transmission mode, for the retrieval of the transmission matrix (left), and with an OCT setup, for the retrieval of the time-gated reflection matrix (right). L: laser source; HWP: half-wave plate; (P)BS: (polarizing) beam-splitter; SLM: liquid-crystal-based spatial light modulator; OBJ: objective lens; TL: tube lens; P: polarizer; M: mirror; CAM: camera.

the whole history of past data, where more contribution is given to newest data. Evidently, in the case of a static scattering medium, all measurements can be equally trusted, thus Eq. (1) reduces to a typical regularized linear least-squares problem upon setting $\lambda = 1$. Once $\lambda$ and $\delta$ are fixed, the least-squares problem has a unique solution, provided the inputs are linearly independent, which is the case in conventional transfer-matrix measurements, where the inputs are drawn from the Hadamard basis of order $N$.

The choice of exponential weights for Eq. (1) is motivated by the physics of our problem. We aim to follow the evolution of the transfer matrix of dynamic scattering media, subjected to uncorrelated variations, whereby the total transferred power fraction is constant in time. These conditions apply in a wide variety of dynamic mechanisms in scattering media investigated with visible and near-infrared light, *e.g.* whenever their inner scatterers move due to functional changes [5, 20], or even when the sample drifts away from its initial position, suggesting that our method can also be used as an online calibration tool of imaging systems. In all these situations, the transfer matrix can indeed be described by the time series [21],

$$\boldsymbol{X}_t = \frac{\sigma_{\boldsymbol{X}}}{\sqrt{\sigma_{\boldsymbol{X}}^2 + \sigma_{\boldsymbol{P}}^2}} (\boldsymbol{X}_{t-1} + \boldsymbol{P}_t) , \tag{2}$$

where we assume that both the transfer matrix and the perturbation matrix $\boldsymbol{P}_t$ are random variables independently drawn from complex Gaussian distributions with zero mean and constant variance $\sigma_{\boldsymbol{X}}^2$ and $\sigma_{\boldsymbol{P}}^2$, respectively [22]. Equation (2) denotes an autoregressive model of order 1, AR(1), whose autocovariance is proportional to $(\sigma_{\boldsymbol{X}}/\sqrt{\sigma_{\boldsymbol{X}}^2 + \sigma_{\boldsymbol{P}}^2})^t$, justifying our exponentially weighted model of Eq. (1). When focusing through dynamic scattering media following Eq. (2), the stability time of the enhancement is proportional to $\sigma_{\boldsymbol{P}}^{-2}$ [21]. This means that the optimal weight $\lambda$ should follow the same dependence, thus in principle requiring the knowledge of the rate of change of the scattering medium. A strategy for automatically tuning the forgetting factor will be discussed in section 4.

Crucially, minimizing the loss function of Eq. (1) does not require storing the whole history of past data. This becomes apparent if we recall that the linear least-squares estimate of $\boldsymbol{X}_t$, $\hat{\boldsymbol{X}}_t$, satisfies the normal equations,

$$\boldsymbol{C}_t \hat{\boldsymbol{X}}_t^H = \boldsymbol{K}_t , \tag{3}$$

with the covariance matrix of inputs and the cross-covariance matrix at time $t$ respectively defined as,

$$\boldsymbol{C}_t \equiv \sum_{\tau=1}^{t} \left( \lambda^{t-\tau} \boldsymbol{a}_\tau \boldsymbol{a}_\tau^H \right) + \delta \lambda^t \boldsymbol{I}_N \in \mathbb{C}^{N \times N} \tag{4a}$$

$$\boldsymbol{K}_t \equiv \sum_{\tau=1}^{t} \lambda^{t-\tau} \boldsymbol{a}_\tau \boldsymbol{y}_\tau^H \in \mathbb{C}^{N \times M} , \tag{4b}$$

with $\boldsymbol{I}_N$ denoting the identity matrix of order $N$ and the superscript $H$ standing for Hermitian transposition. The quantities calculated in Eqs. (4) can be both estimated recursively, as follows:

$$\boldsymbol{C}_t = \lambda \boldsymbol{C}_{t-1} + \boldsymbol{a}_t \boldsymbol{a}_t^H \tag{5a}$$

$$\boldsymbol{K}_t = \lambda \boldsymbol{K}_{t-1} + \boldsymbol{a}_t \boldsymbol{y}_t^H . \tag{5b}$$

Equations (5) mean the loss defined in Eq. (1) can be minimized from the new piece of data $(\boldsymbol{a}_t, \boldsymbol{y}_t)$ and the previous estimates of the covariance and cross-covariance matrices, whose sizes are independent of the amount of past data. It becomes now clear how the RLS algorithm combines the benefits of transfer-matrix-based and optimization approaches. Using a recursive procedure, a typical asset of, *e.g.*, the continuous sequential algorithm (CSA), the partitioning algorithm [21], or more computationally intense genetic algorithms [15], the full $\boldsymbol{X}_t$ is estimated *in parallel* at all output pixels, thereby preserving all light-control capabilities allowed by the knowledge of the transfer matrix [10–12, 14] [Fig. 1(b)]. In principle, the transfer matrix could be obtained from Eq. (3) as $\hat{\boldsymbol{X}}_t = \boldsymbol{K}_t^H (\boldsymbol{C}_t^{-1})^H$. However, in what follows we will implement the *inverse QR-decomposition-based RLS* (abbreviated as inverse QRD-RLS) algorithm [23]. Because it avoids matrix inversions and it always preserves the non-negativeness of the covariance matrix, it possesses higher numerical stability than directly inverting Eq. (3). Overall, it boils down to performing a QR decomposition of a matrix constructed from the new data and the previous estimate of the square root of the inverse covariance matrix. This results in few lines of code which can be readily implemented in any programming language using standard libraries or built-in functions (see the box Algorithm 1 and the corresponding code available at Ref. [24]). As can be seen from Eq. (1) and Algorithm 1, the regularization constant $\delta$ is used to construct the initial estimate of the square root of the inverse correlation matrix, hence it mostly impacts the convergence speed at early iterations. In section 4, the choice of its value will be discussed.

---

**Algorithm 1:** Inverse QRD-RLS update

Initializations: $\hat{\boldsymbol{X}}_0 = \mathbf{0}$, $(\boldsymbol{C}_0^{-1})^{1/2} = \delta^{-1/2}\boldsymbol{I}_N$

---

**Input:** New input pattern $\boldsymbol{a}_t$, new output pattern $\boldsymbol{y}_t$, previous estimate of the transfer matrix $\hat{\boldsymbol{X}}_{t-1}$, previous estimate of the square root of the inverse covariance matrix $(\boldsymbol{C}_{t-1}^{-1})^{1/2}$, forgetting factor $\lambda$

`/* Construction of the matrix` $\boldsymbol{U}$ `*/`

$$\boldsymbol{U} = \begin{bmatrix} 1 & \lambda^{-1/2}\boldsymbol{a}_t^H(\boldsymbol{C}_{t-1}^{-1})^{1/2} \\ \mathbf{0} & \lambda^{-1/2}(\boldsymbol{C}_{t-1}^{-1})^{1/2} \end{bmatrix}$$

`/* QR decomposition of` $\boldsymbol{U}^H$ `*/`

$$\boldsymbol{U}^H = \boldsymbol{Q}\boldsymbol{V}^H$$

$$\boldsymbol{V} = \begin{bmatrix} v_{11} & \mathbf{0}^H \\ \boldsymbol{v}_{21} & (\boldsymbol{C}_t^{-1})^{1/2} \end{bmatrix}$$

`/* Update of the transfer matrix */`

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_{t-1} + (\boldsymbol{y}_t - \hat{\boldsymbol{X}}_{t-1}\boldsymbol{a}_t)\boldsymbol{v}_{21}^H v_{11}^{-1}$$

**return** $\hat{\boldsymbol{X}}_t$ and $(\boldsymbol{C}_t^{-1})^{1/2}$

---

## 3. Experiments

Figure 1(c) shows the sketches of the experimental implementations used to demonstrate our method. Both are based on phase-shifting digital holography to retrieve the complex output fields $\boldsymbol{y}_t$ after interacting with a multiply scattering medium. The medium is an opaque deposit of ZnO nanoparticles (size < 100 nm, relative transmittance ~0.15), whose thickness (20 µm) is 5 to 7 transport mean free paths, ensuring full mixing of its optical modes at the output. The input fields are shaped via a reflective, phase-only and liquid-crystal-based spatial light modulator (SLM, Meadowlark Optics HSP512L-1064) and focused on the scattering medium with an objective with a numerical aperture of 0.4 (Olympus PLN20X). A region-of-interest containing ~80 speckle grains is imaged onto a CCD camera (Manta G-046B, Allied Vision) via a tube lens, yielding a pixel size of 0.2 µm at the CCD plane. Before impinging onto the SLM, part of the beam is redirected along a reference arm with a polarizing beam splitter (PBS), and subsequently recombined with the scattered beam through a beam splitter (BS). The relative power of the two beams, yielding the maximum interference contrast, is adjusted via two half-wave plates, one along the common path and one along the reference arm, while a polarizer in front of the camera filters out any potential residual ballistic component traveling along with the scattered beam. In the experiments in transmission [Fig. 1(c), left], the beam exiting the scattering medium is collected at a distance of ~1.5 mm, where a fully developed speckle pattern was observed, with another Olympus PLN20X 0.4 NA objective. The light source (MaiTai HP Ti:Sapphire laser, Spectra-Physics) is set to monochromatic operation mode at a wavelength of 808 nm. The experiments in reflection [Fig. 1(c), right] reproduce a typical optical coherence tomography

(OCT) setup, whereby ultrashort pulses (with a central wavelength of 808 nm and a duration of 100 fs) are sent through the scattering medium and the backscattered, elongated pulses are gated at a time delay set by a delay line along the reference arm.

Dynamics is introduced by transversally translating the scattering medium across the incident beam, with independent and randomly distributed Gaussian steps, whose standard deviation determines the stability time of the medium. More details on it will be provided in the next section.

## 4. Results

Figure 2 summarizes the performance of the RLS algorithm for the online estimation of the transmission matrix. The beam incident onto the SLM is modulated according to the Hadamard patterns with $N = 64$ pixels. Every time an input $\boldsymbol{a}_t$ is sent through the scattering medium and the corresponding output field $\boldsymbol{y}_t$ is measured, the inverse QRD-RLS update routine of Algorithm 1 is executed, yielding an estimate $\hat{\boldsymbol{X}}_t$ of the transfer matrix. Note, that this procedure can be continuously repeated—after sending the $N$-th input, the first Hadamard vector or any other known input pattern can be sent. As long as the scattering medium is static, probing it with the same input multiple times corresponds to oversampling the unknown $N \times M$ coefficients of its transfer matrix, thereby improving their estimation. It is indeed known that the covariance of the estimated transfer matrix is inversely proportional to $\boldsymbol{C}_t^{-1}$, thus decreasing as $t^{-1}$ [17]. Since the true value $\boldsymbol{X}_t$ is unknown, the quality of our reconstruction is evaluated via the intensity of a focus produced behind the scattering medium. We report the intensity enhancement, relative to the average intensity of a non-optimized speckle pattern [8]. The learning curve for a static scattering medium, obtained from the RLS algorithm, is shown as an orange trace in Fig. 2(a). The temporal axis is expressed in units of $T_{TM}$, which is defined as the time needed to update the estimation of transfer matrix $N$ times. In other words, a conventional transfer matrix experiment lasts $T_{TM}$. Equivalently, a normalized time of 2 means the oversampling ratio is 2. To showcase the beneficial effect of oversampling, the blue trace shows the performance of a conventional transfer-matrix measurement, lasting until $t/T_{TM} = 1$, thus using $N$ measurements. At times $t/T_{TM} \leq 1$, the two approaches are equivalent—data are not oversampled. At later times, however, one can take advantage of the whole history of past data to build an estimate more resilient against noise. Our values of the enhancement, when compared to the number of input degrees of freedom $N$, are on a par with previously reported measurements with no oversampling [13, 25, 26].

The same procedure is repeated with dynamic scattering media. By duly tuning the amplitude of their movements, we achieve different stability times $T_{stab}$ (also expressed in units of $T_{TM}$). These are estimated as the time constant of an exponential function fitting the tails of the blue traces. Note, that the scattering medium is dynamic for the whole duration of the experiments. At oversampling ratios in the range 3-4, we increase the focus intensity by a factor between 1.5 and 2, compared to the values after a conventional transfer-matrix approach. Upon decreasing $T_{stab}$, the oversampling ratio decreases too, and the performances of the two approaches gradually match, however the RLS estimation always operates in a memory-efficient manner. With dynamic media, forgetting factors $\lambda < 1$ should be used. In our experiments featuring $T_{stab}$ in the range 1-4, we have chosen $1 - \lambda \approx 10^{-5}$, achieving a good compromise between tracking capability and numerical stability. Interestingly, it has been shown that the optimal forgetting factor heavily depends on the number of unknown parameters $N$ which, fortunately, is under user control [27]. Furthermore, the structure of Algorithm 1 suggests that each inverse QRD-RLS iteration may be run at a different value of $\lambda$, allowing the user to pick the one yielding the best performance in an online manner, *i.e.*, with no need to restart the optimization anew and using the current enhancement as a feedback to tune the next value of $\lambda$. Trivially, the optimal value for static media is instead $\lambda = 1$. The best regularization constant $\delta$ depends on the SNR of the measurements. In our experiments, the fact that the RLS algorithm is on a par with the conventional transfer-matrix approach at $t/T_{TM} \leq 1$ and $T_{TM} < T_{stab}$ suggests that the selected regularization constant (here $\delta = 1$) is optimized for the best performance.

In order to test the suitability of the cost function in Eq. (1) and the model in Eq. (2) to our experimental settings, the experiments at the top row of Fig. 2 are reproduced with numerical simulations, following the AR(1) model of Eq. (2) [Figs. 2(d)-(f)]. We simulated a finite SNR by corrupting the outputs $\boldsymbol{y}_t$, with average power $\sigma_{\boldsymbol{y}}^2$, with additive white complex Gaussian noise with variance $\sigma_{\text{noise}}^2$ and setting the simulated SNR as $\text{SNR}_{\text{sim}} \equiv \sigma_{\boldsymbol{y}}^2/\sigma_{\text{noise}}^2$. In experiments, the noise was estimated from the standard deviation of the intensity enhancement at $t/T_{TM} > 1$ in Fig. 2(a), while the signal level was calculated as the average intensity of *non-optimized* speckle patterns, with the ratio between the two yielding the experimental SNR, $\text{SNR}_{\text{exp}}$. At comparable values of stability times and $\text{SNR}_{\text{exp}}$ and $\text{SNR}_{\text{sim}}$ [noted in Figs. 2(a) and (d), respectively, for a static medium], quantitative agreement is overall obtained. For example, if the SNR is increased by a factor of 2 doubling the number of phase-stepped images for field reconstruction, the experimental performance is the one plotted as an inset in Fig. 2(a). The same trend is retrieved by simulating measurements with halved $\sigma_{\text{noise}}^2$ [inset in Fig. 2(d)].

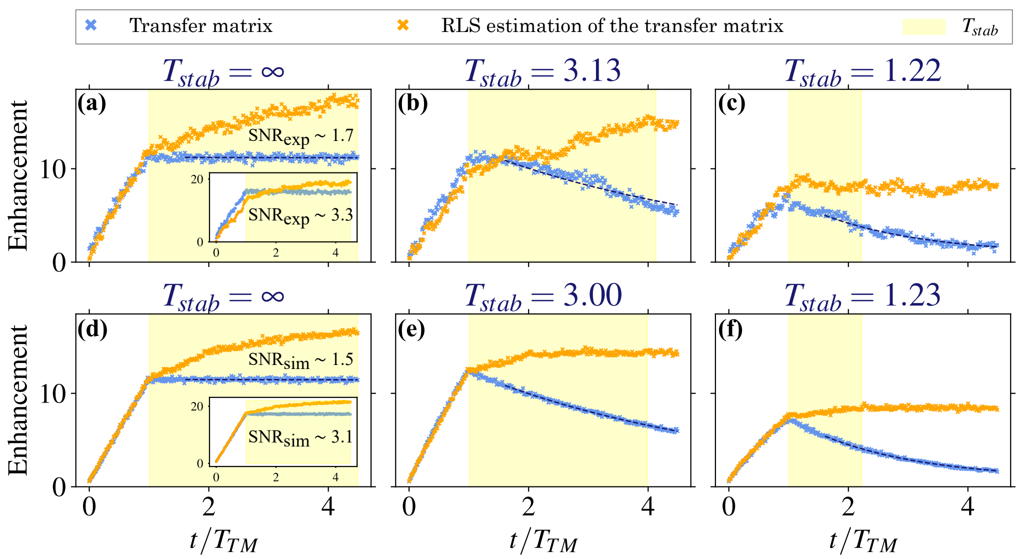


Fig. 2. Enhancement of the intensity at one output pixel produced through a scattering medium, as a function time. $T_{TM}$ is the time to optimize over all the $N = 64$ degrees of freedom once. Blue points: conventional transfer-matrix measurement, lasting until $t/T_{TM} = 1$. Its estimate is held constant for later times, allowing the extraction of the stability time $T_{stab}$ of the scattering medium via exponential fitting (dashed dark blue traces). Note that, when a finite $T_{stab}$ is reported, the scattering medium is dynamic for the whole duration of the experiment. Orange points: RLS estimation of the transfer matrix. (a)-(c): experimental results, averaged over 9 realizations of a focus produced at the center of the camera field of view, upon measuring different regions of the scattering medium. (d)-(f): corresponding simulations at comparable $T_{stab}$. The insets in panels (a), (d) show the same results obtained after doubling the SNR, yielding an improvement of the enhancement by a factor of $\sim \sqrt{2}$. The SNR is noted in panels (a) and (d) for the case of a static medium.

Analogous results, plotted in Fig. 3, are obtained with the non-invasive OCT setup on the right-hand side of Fig. 1(c), setting the time delay yielding the maximum average gated intensity. In this instance, the transfer matrix is the time-gated reflection matrix [28]. The two learning curves corresponding to the retrieval of the transfer matrix are compared to a conventional optimization routine, which can recursively track the changes in the scattering medium, namely the CSA (in cyan). After blocking the beam along the reference arm, we implement a version of the CSA modulating half of the SLM pixels (corresponding to the +1 or -1 entries of the $N$ Hadamard patterns) at each iteration, yielding the best interference contrast (thus bearing similarities to the partitioning algorithm too [21]). It displays comparable performances to a conventional transfer-matrix measurement with a static medium [although convergence is reached later, owing to its stochastic nature, Fig. 3(a)], and it shows solid tracking capabilities in dynamic environments [Figs. 3(b)-(c)]. Still, the resilience to noise of the inverse QRD-RLS algorithm makes it the preferred choice in this setting too, achieving an intensity twice as high as the one obtained with the CSA. The bottom row of Fig. 3 shows the corresponding focal spots produced by each algorithm at the last time step.

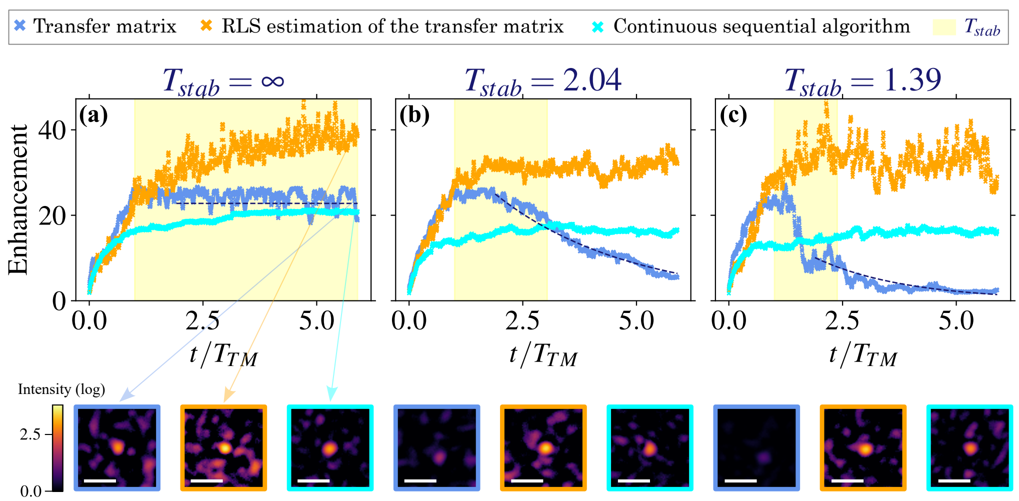


Fig. 3. Enhancement of the intensity at one output pixel produced in a scattering medium, as a function time. $T_{TM}$ is the time to optimize over all the $N = 256$ degrees of freedom once. Blue points: conventional transfer-matrix measurement, lasting until $t/T_{TM} = 1$. Its estimate is held constant for later times, allowing the extraction of the stability time $T_{stab}$ of the scattering medium via exponential fitting (dashed dark blue traces). When a finite $T_{stab}$ is reported, the scattering medium is dynamic for the whole duration of the experiment. Orange points: RLS estimation of the transfer matrix. Cyan points: optimization via the continuous sequential algorithm (CSA). (a)-(c): experimental tracking performance, averaged over 8 realizations of foci produced across the full camera field of view, whence the higher variability than in Fig. 2. The bottom row displays typical images of the focus at the last time step, for each algorithm and each value of $T_{stab}$ (in a common logarithmic scale to ease visibility and comparison). All the images show one realization only, and not an average of the 8 realizations. Note that the foci produced by the transfer matrix and RLS algorithms stand on a higher speckle background than those obtained from the CSA because of the superposition with the static reference field needed for interferometric phase estimation. We have accordingly accounted for this aspect when estimating the enhancements reported in the top row. The length of the white scale bars is 5 µm. The data underlying this figure are available in Dataset 1 at Ref. [29].

Because our proposed routine retrieves the coefficients of the transfer matrix at all output pixels simultaneously, its applications go beyond focusing. Figure 4 showcases two light-control tasks, through dynamic scattering media, enabled by the recursive and online estimation of the transfer matrix. The first one is maximal energy transmission, upon sending the leading singular vector of the transfer matrix (top row) [12], and the second one consists in arbitrarily shaping a PSF (bottom row), here in a donut shape [10]. As expected, these trends replicate the performance on the focusing task of Fig. 2.

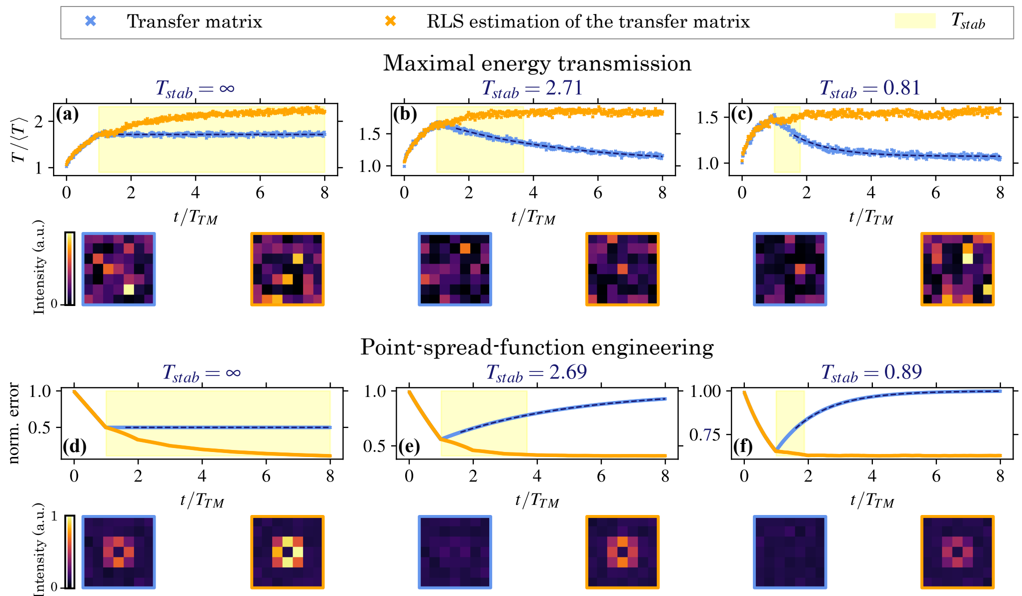


Fig. 4. Light control through dynamic scattering media goes beyond focusing. (a)-(c): Enhancement of the total transmittance across the full field of view through a scattering medium, as a function time and for different stability times $T_{stab}$. $T_{TM}$ is the time to optimize over all the $N$ degrees of freedom once. Blue points: conventional transfer-matrix measurement, lasting until $t/T_{TM} = 1$. Its estimate is held constant for later times, allowing the extraction of the stability time $T_{stab}$ of the scattering medium via exponential fitting (dashed dark blue traces). When a finite $T_{stab}$ is reported, the scattering medium is dynamic for the whole duration of the experiment. Orange points: RLS estimation of the transfer matrix. Here we used $N = 64$ and $M = 49$. (d)-(f): normalized error in the estimation of the transfer matrix of the scattering medium, used to project a donut-shaped beam at the output plane. The same color legend as in (a)-(c) applies. In each set of results, the bottom row displays typical camera images at the last time step. Simulated results.

The most important asset to characterize dynamic scattering media is the wavefront shaping device. Digital micromirror devices (DMD) offer a valuable alternative to liquid-crystal-based arrays and microelectromechanical systems (MEMS) modulators in terms of cost (~1 kUSD), pixel count (>$10^5$) and operating frequencies (>10 kHz). Despite their binary amplitude modulation, several strategies have been devised to enable light control through scattering media. Lee holography [30, 31] and superpixel-based related methods [32] achieve phase and amplitude control, at the expense of a more involved setup and a relatively low light efficiency, as they rely on an analog spatial filter. Aiming for a simple and non-invasive implementation suitable for real-life applications, Bayesian algorithms have been proposed to solve the phase retrieval problem $\boldsymbol{y}_t = |\boldsymbol{X}_t \boldsymbol{a}_t|^2$ (with $|\cdot|$ denoting the element-wise modulus operation), *i.e.*, recover the transfer matrix from intensity-only measurements and binary amplitude modulation of the inputs, thus transferring the hardware complexity to the software. Examples include the phase retrieval Variational Bayes Expectation-Maximization (prVBEM) [33, 34] algorithm, the phase retrieval Swept Approximate Message Passing (prSAMP) [35] algorithms, and their corresponding compressive version, named phase retrieval Generalized AMP (prGAMP) [36]. However, their complexities are of order $O(t^2)$ per iteration, preventing their application to real-time online learning of long ($t \gg N$) and multidimensional time series.

In what follows, we show how to implement the RLS estimation technique using non-invasive, intensity-only measurements and binary amplitude modulation of the inputs. When performing wavefront shaping experiments with a DMD, light control is restricted to opening or blocking the modes of the scattering medium, so to achieve the desired output patterns. Hence, the knowledge of the complex-valued transfer matrix is of limited use. We now build on the contribution by Tao and colleagues [37]. They regard each binary input $\boldsymbol{a}_t \in \{0,1\}^N$ as the sum of the first Hadamard vector $\boldsymbol{h}_1 = \{1\}^N$, referred to as "reference", with any other Hadamard vector $\boldsymbol{h}_t \in \{+1,-1\}^N$, namely $\boldsymbol{a}_t = (\boldsymbol{h}_1 + \boldsymbol{h}_t)/2$. In a similar fashion to inline digital holography, in the output pixels where the reference intensity is larger than the response to an average input, the phase retrieval equation can be linearized. They derive the following linear approximation,

$$\frac{1}{2}|\boldsymbol{X}_t\boldsymbol{h}_1| \circ \left(|\boldsymbol{X}_t\boldsymbol{a}_t|^2 \oslash |\boldsymbol{X}_t\boldsymbol{h}_1|^2 - \mathbf{1}\right) \approx \mathrm{Re}\{\boldsymbol{X}_t\}\boldsymbol{h}_t\,, \tag{6}$$

where $\circ$ and $\oslash$ denote element-wise vector multiplication and division, respectively, $\mathbf{1} \equiv \{1\}^M$ and $\mathrm{Re}\{\cdot\}$ stands for real part. Note, that the condition for a proper linearization is met, assuming the output pixels are independent, with a probability

$$\mathcal{P}(I > \langle I\rangle) = \int_{\langle I\rangle}^{\infty} p(I)dI = e^{-1} \approx 40\%\,, \tag{7}$$

where we have used the probability distribution of the speckle intensity, $p(I) \equiv \exp\left(-I/\langle I\rangle\right)/\langle I\rangle$ [38]. As all the terms in its left-hand side $\tilde{\boldsymbol{y}}_t$ are known, we can recursively solve Eq. (6) for $\mathrm{Re}\{\boldsymbol{X}_t\}$, minimizing a loss function like the one in Eq. (1), and interpreting $\boldsymbol{h}_t$ and $\tilde{\boldsymbol{y}}_t$ as real inputs and outputs, respectively. The real (or, equivalently, imaginary) part of the transfer matrix is all is needed to focus at any output pixel where the linear approximation holds. The corresponding results in Fig. 5 indeed show the same trend as in Figs. 2, 3 and 4. Here, the enhancement is expressed relative to the maximum enhancement achievable with binary amplitude modulation $\approx 1 + (N/2-1)/\pi$ [39]. Feedback-based routines, like the binary version of the CSA [39] (plotted in cyan in Fig. 5), are highly impacted by experimental noise, as they rely on one single output value. In contrast, exploiting the past data allows us to provide solutions more resilient to noise.

To gain more insight into the performance of our experimental system, its throughput is estimated with the parameters from [13], namely $M = 256$ and 4 phase-shifted intensity images to evaluate each output field. In Fig. 6 we plot, as a function of the number of input modes $N$, the time to update the optimal focusing pattern from one new piece of data, therefore comprising one (complex) output measurement, the update of the transfer matrix and the computation of the optimal input pattern. For a sufficiently low number of input modes ($N \leq 256$ in our implementation), the bottleneck is set by the refresh rate of the SLM—we indeed recover a baseline at ~50 ms, which is consistent with the response time $\gtrsim$ 10 ms reported by the manufacturer. With increasing values of $N$, the update of the transfer matrix and the computation of the optimal pattern take a non-negligible time at each iteration, hence an onset at $N \sim 256$ is observed. In a conventional transfer-matrix measurement (blue line and data points) performed with Hadamard inputs, an additional $O(N^2)$ is required to bring the optimal focusing pattern from the Hadamard to the canonical basis. The inverse QRD-RLS estimation technique (orange line and data points), based on Algorithm 1, would run with a $O(N^3)$ complexity, as it involves a QR decomposition [41], but we retrieve a lower power dependence (~2.6) owing to the low number of data points above the onset. We should, however, recall that Algorithm 1 has been implemented to enjoy superior numerical stability. A typical RLS algorithm propagating the

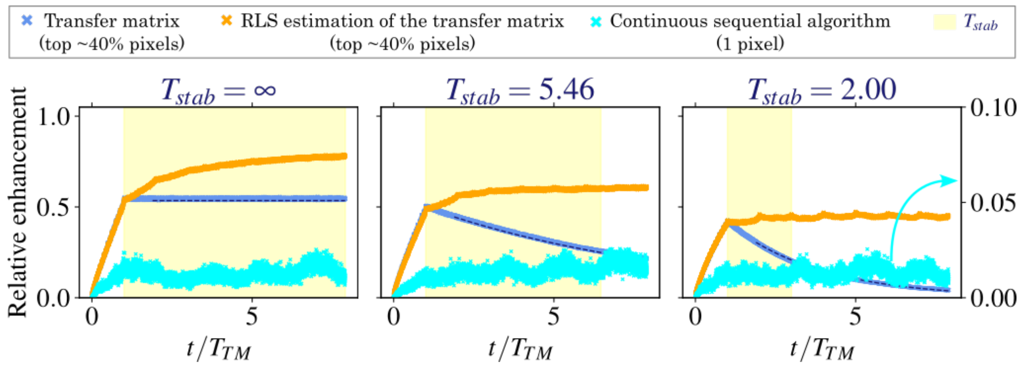


Fig. 5. Enhancement of the intensity produced at one output pixel through a scattering medium, as a function time and for different stability times $T_{stab}$. By using the linear approximation of Eq. (6) (valid across ~40% of the output pixels), we achieve wavefront shaping from intensity-only images. $T_{TM}$ is the time to optimize over all the $N$ degrees of freedom once. Blue points: conventional transfer-matrix measurement, lasting until $t/T_{TM} = 1$. Its estimate is held constant for later times, allowing the extraction of the stability time $T_{stab}$ of the scattering medium via exponential fitting (dashed dark blue traces). When a finite $T_{stab}$ is reported, the scattering medium is dynamic for the whole duration of the experiment. Orange points: RLS estimation of the transfer matrix. Cyan points: continuous sequential algorithm (CSA), plotted on a different scale on the right-hand vertical axis. To reproduce noisy measurements with suboptimal detector performance, all simulated intensities $I$ were corrupted with additive Gaussian noise with a standard deviation of $20\sqrt{I}$ [40]. Simulated results.

inverse covariance matrix instead of its square root would require $N^2$ operations, thus matching the performance of a conventional transfer-matrix measurement. Owing to its updating routine, the iteration time of the CSA (cyan line and data points) is not impacted by the number of input modes, however its performance is limited in dynamic and noisy environments as shown above. As a final remark we stress that online optimization is run on the CPU of an Intel Core i7-6700 processor with 4 cores, a clock speed of 3.4 GHz and 16 GB RAM, thus yielding the onset at $N \sim 256$. Therefore our experiments optimize over $N \leq 256$ modes. Such figures do not represent a bottleneck for real-time and online wavefront shaping at high enhancements. The number $N$ can be definitely increased on a high-performance computing platform, for example implementing the RLS algorithm on a FPGA [42] (as was already done in [25] with a conventional continuous optimization algorithm for focusing through dynamic scattering media), or on a GPU [43].

## 5. Outlook

We have presented a recursive and online optimization procedure for the estimation of the transfer matrix of dynamic scattering media, combining the benefits of optimization-based routines and transfer-matrix measurements in wavefront shaping. Experimental and numerical demonstrations have been provided on conventional wavefront shaping setups and for different light-control tasks, noise levels and stability times. Its most intriguing feature is the possibility to optimize multi- and high-dimensional transfer matrices, without the need to store the history of past data in memory. Therefore, we foresee our method to turn out pivotal whenever the scattering behaviour of living biological specimens has to be tracked at various timescales.

In our proof-of-principle experiments, all optical modes change with the same rate, therefore they share the same oversampling ratio. However, when imaging large fields of view ($\sim 10^4 \mu m^2$)

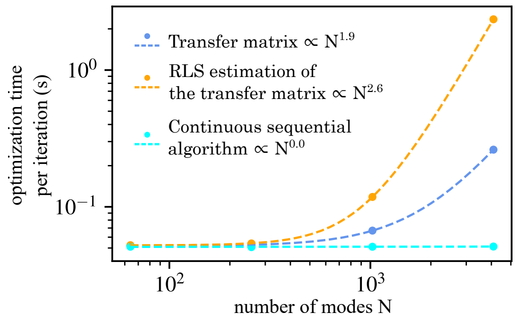


Fig. 6. Time required to update the optimal focusing pattern from one new piece of data, estimated in the experimental setup of Fig. 1(c) right, as a function of the number of input modes. Four phase-stepped intensity images are combined to estimate each output field across a field of view of $M = 256$ pixels. Blue points and line: conventional transfer-matrix measurement; orange: RLS estimation of the transfer matrix; cyan: continuous sequential algorithm (CSA). Each point is an average of 4096 measurements, such that the standard deviation of the mean is always within the marker size.

in biological media, timescales differing by factors as large as 100 are accessible. For example, the modes induced by blood flowing decorrelate in less than 10 ms (>100 Hz), while breathing modes can last as long as 800 ms (1.25 Hz) in mice [5]. As as result of that, the slowest modes enjoy an oversampling ratio close to 100. This means that, compared to an offline least-square estimation of the transfer matrix, a factor of 100 is saved in memory, which can be ultimately used to enlarge the field of view by 2 orders of magnitude. Using the latest MEMS modulators, $N =$ 600 modes can be optimized at a rate of 60 kHz in 10 ms, thus allowing the transfer matrix to be estimated at $M \sim 1.6 \cdot 10^6$ output pixels in parallel, assuming 16 GB RAM and double-precision floating-point format (16 B per complex matrix element). This is illustrated in Fig. 7(a), where the feasibility region for offline least-squares is shaded in blue and depends on the oversampling ratio. On the other hand, using the RLS estimation means oversampling does not play a role, so its feasibility region is much larger (orange shaded area). If we also consider that, in ultrafast wavefront shaping systems like [25], the SNR approaches 1, at an oversampling ratio of 100 the RLS estimation of the transfer matrix yields an improvement of the focus intensity by a factor of 2, compared to a conventional transfer-matrix measurement with no oversampling [Fig. 7(b)]. Focusing deep inside scattering media, at locations characterized by a specific stability time, may become a reality, thanks to the recent advancements in optimal light control, exploiting the knowledge of the transfer matrix measured at different times [44, 45].

Besides sharing the same stability times, all the optical modes considered here are also unpredictable, as they feature random and independent increments according to Eq. (2). Should one possess prior knowledge on the medium dynamics, the RLS estimation may even be employed to predict future scattering behaviours as well as informing the user on the next most informative inputs to optimize information retrieval [46]. For example, breathing modes and heartbeat are known to induce revivals of correlations [5]. Such behaviour was neither reproduced in our experiments, nor accounted for in our model. In this context, the investigation of dynamic biological tissues would benefit from an implementation of the RLS algorithm dealing with quasi-periodic measurements.

We would finally like to remind that the effectiveness of the RLS algorithm is enabled by the linear relationship between the input and output patterns. Linearity is guaranteed by light-matter interaction via elastic scattering and thanks to our measurement scheme, allowing quantitative phase estimation of the output fields. When a non-linear transfer function was involved (Fig. 5), a linear approximation was made, at the cost of reduced performance. Towards the recursive optimization of non-linear functions, a kernel version of the RLS algorithm has been proposed [47,48]. It relies on performing linear regressions in a higher ($> N$) dimensional feature space, approximating the non-linear function. Its implementation, although more complicated than its linear counterpart, would be worth investigating, as it would unlock online learning of the transfer matrix of dynamic scattering media for a wide variety of contrast mechanisms, from fluorescence to non-linear coherent scattering.

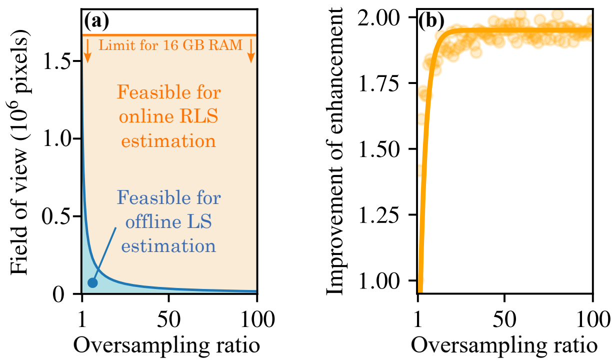


Fig. 7. (a) Feasibility regions of the offline least-squares algorithm (blue shaded area) and of the RLS algorithm (orange shaded area), on the plane spanned by the oversampling ratio and the field of view. (b) Evolution of the enhancement as a function of the oversampling ratio (relative to the value at oversampling = 1) at SNR = 1. Points: simulations; line: exponential fit.

**Funding.** Swiss National Science Foundation (project P400P2_199329); H2020 European Research Council (SMARTIES-724473); Horizon 2020 Framework Programme (863203).

**Acknowledgments.** We wish to thank Fabrice Harms (Imagine Optic) for drawing our attention to RLS algorithms and Bernhard Rauer (LKB), Louisiane Devaud (LKB) and Jonathan Dong (EPFL) for contributing insightful comments.

**Disclosures.** The authors declare no conflicts of interest.

**Data Availability Statement.** Reconstruction and simulation codes underlying the results presented in this paper are available in Ref. [24]. The data underlying Fig. 3 are available in Dataset 1 at Ref. [29].

## References

1. J. Mertz, *Introduction to optical microscopy* (Cambridge University Press, 2019).
2. G. C. Ellis-Davies, "Caged compounds: photorelease technology for control of cellular chemistry and physiology," Nat. Methods **4**, 619–628 (2007).
3. J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," Nat. Methods **2**, 910–919 (2005).
4. S. Rotter and S. Gigan, "Light fields in complex media: Mesoscopic scattering meets wave control," Rev. Mod. Phys. **89**, 015005 (2017).
5. Y. Liu, P. Lai, C. Ma, X. Xu, A. A. Grabar, and L. V. Wang, "Optical focusing deep inside dynamic scattering media with near-infrared time-reversed ultrasonically encoded (TRUE) light," Nat. Commun. **6**, 1–9 (2015).
6. H. E. Assmus, R. Herwig, K.-H. Cho, and O. Wolkenhauer, "Dynamics of biological systems: role of systems biology in medical research," Expert Rev. Mol. Diagn. **6**, 891–902 (2006).
7. H. W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster, "Inverse problems in systems biology," Inverse Probl. **25**, 123014 (2009).
8. I. M. Vellekoop, "Feedback-based wavefront shaping," Opt. Express **23**, 12189–12206 (2015).
9. L. Devaud, B. Rauer, J. Melchard, M. Kühmayer, S. Rotter, and S. Gigan, "Speckle engineering through singular value decomposition of the transmission matrix," Phys. Rev. Lett. **127**, 093903 (2021).
10. A. Boniface, M. Mounaix, B. Blochet, R. Piestun, and S. Gigan, "Transmission-matrix-based point-spread-function engineering through a complex medium," Optica **4**, 54–59 (2017).
11. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," Nat. Commun. **1**, 81 (2010).

12. M. Kim, Y. Choi, C. Yoon, W. Choi, J. Kim, Q.-H. Park, and W. Choi, "Maximal energy transport through disordered media with the implementation of transmission eigenchannels," Nat. Photonics **6**, 581–585 (2012).
13. S. Popoff, G. Lerosey, R. Carminati, M. Fink, A. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," Phys. Rev. Lett. **104**, 100601 (2010).
14. M. W. Matthès, Y. Bromberg, J. de Rosny, and S. M. Popoff, "Learning and avoiding disorder in multimode fibers," Phys. Rev. X **11**, 021060 (2021).
15. D. B. Conkey, A. N. Brown, A. M. Caravaca-Aguirre, and R. Piestun, "Genetic algorithm optimization for focusing through turbid media in noisy environments," Opt. Express **20**, 4840–4849 (2012).
16. S. Cheng, T. Zhong, C. M. Woo, Q. Zhao, H. Hui, and P. Lai, "Long-distance pattern projection through an unfixed multimode fiber with natural evolution strategy-based wavefront shaping," Opt. Express **30**, 32565–32576 (2022).
17. S. S. Haykin, *Adaptive filter theory* (Pearson Education, 2014).
18. S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," IEEE Trans. Signal Process. **45**, 1118–1128 (1997).
19. A. Patra, S. Das, S. Mishra, and M. R. Senapati, "An adaptive local linear optimized radial basis functional neural network model for financial time series prediction," Neural. Comput. Appl. **28**, 101–110 (2017).
20. J. Brake, M. Jang, and C. Yang, "Analyzing the relationship between decorrelation time and tissue thickness in acute rat brain slices using multispeckle diffusing wave spectroscopy," J. Opt. Soc. Am. A **33**, 270–275 (2016).
21. I. M. Vellekoop and A. Mosk, "Phase control algorithms for focusing light through turbid media," Opt. Commun. **281**, 3071–3080 (2008).
22. M. Webster, T. Gerke, A. Weiner, and K. Webb, "Spectral and temporal speckle field measurements of a random medium," Opt. Lett. **29**, 1491–1493 (2004).
23. S. T. Alexander and A. L. Ghimikar, "A method for recursive least squares filtering based upon an inverse QR decomposition," IEEE Trans. Signal Process. **41**, 20 (1993).
24. L. Valzania, "Online learning of the transfer matrix of dynamic scattering media," `https://github.com/laboGigan/online_learning_TM` (2022).
25. B. Blochet, L. Bourdieu, and S. Gigan, "Focusing light through dynamical samples using fast continuous wavefront optimization," Opt. Lett. **42**, 4994–4997 (2017).
26. B. Blochet, K. Joaquina, L. Blum, L. Bourdieu, and S. Gigan, "Enhanced stability of the focus obtained by wavefront optimization in dynamical scattering media," Optica **6**, 1554–1561 (2019).
27. S. Ciochina, C. Paleologu, J. Benesty, and A. A. Enescu, "On the influence of the forgetting factor of the RLS adaptive filter in system identification," in *2009 International Symposium on Signals, Circuits and Systems,* (IEEE, 2009), pp. 1–4.
28. A. Badon, D. Li, G. Lerosey, A. C. Boccara, M. Fink, and A. Aubry, "Smart optical coherence tomography for ultra-deep imaging through highly scattering media," Sci. Adv. **2**, e1600370 (2016).
29. L. Valzania, "Dataset for 'Online learning of the transmission matrix of dynamic scattering media: wavefront shaping meets multidimensional time series'," `figshare_url_to_be_included_in_proof` (2022).
30. W.-H. Lee, "Binary synthetic holograms," Appl. Opt. **13**, 1677–1682 (1974).
31. D. B. Conkey, A. M. Caravaca-Aguirre, and R. Piestun, "High-speed scattering medium characterization with application to focusing light through turbid media," Opt. Express **20**, 1733–1740 (2012).
32. S. A. Goorden, J. Bertolotti, and A. P. Mosk, "Superpixel-based spatial amplitude and phase modulation using a digital micromirror device," Opt. Express **22**, 17999–18009 (2014).
33. A. Drémeau, A. Liutkus, D. Martina, O. Katz, C. Schülke, F. Krzakala, S. Gigan, and L. Daudet, "Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques," Opt. Express **23**, 11898–11911 (2015).
34. A. Drémeau and F. Krzakala, "Phase recovery from a Bayesian point of view: the variational approach," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* (IEEE, 2015), pp. 3661–3665.
35. B. Rajaei, E. W. Tramel, S. Gigan, F. Krzakala, and L. Daudet, "Intensity-only optical compressive imaging using a multiply scattering material and a double phase retrieval approach," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* (IEEE, 2016), pp. 4054–4058.
36. C. A. Metzler, A. Maleki, and R. G. Baraniuk, "BM3D-PRGAMP: Compressive phase retrieval based on BM3D denoising," in *2016 IEEE International Conference on Image Processing (ICIP),* (IEEE, 2016), pp. 2504–2508.
37. X. Tao, D. Bodington, M. Reinig, and J. Kubby, "High-speed scanning interferometric focusing by fast measurement of binary transmission matrix for channel demixing," Opt. Express **23**, 14168–14187 (2015).
38. J. W. Goodman, *Speckle phenomena in optics: theory and applications* (Roberts and Company Publishers, 2007).
39. D. Akbulut, T. J. Huisman, E. G. van Putten, W. L. Vos, and A. P. Mosk, "Focusing light through random photonic media by binary amplitude modulation," Opt. Express **19**, 4017–4029 (2011).
40. L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, "Experimental robustness of Fourier ptychography phase retrieval algorithms," Opt. Express **23**, 33214–33240 (2015).
41. L. N. Trefethen and D. Bau III, *Numerical linear algebra*, vol. 50 (SIAM, 1997).
42. M. Karkooti, J. R. Cavallaro, and C. Dick, "Fpga implementation of matrix inversion using qrd-rls algorithm," in *Asilomar Conference on Signals, Systems, and Computers,* (2005).
43. Y. Dua, V. Kumar, and R. S. Singh, "Parallel lossless hsi compression based on rls filter," J. Parallel Distrib. Comput. **150**, 60–68 (2021).

44. D. Bouchet, S. Rotter, and A. P. Mosk, "Maximum information states for coherent scattering measurements," Nat. Phys. **17**, 564–568 (2021).
45. D. Bouchet, L. M. Rachbauer, S. Rotter, A. P. Mosk, and E. Bossy, "Optimal control of coherent light scattering for binary decision problems," Phys. Rev. Lett. **127**, 253902 (2021).
46. M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, "Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction," Phys. Rev. X **10**, 041037 (2020).
47. Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," IEEE Trans. Signal Process. **52**, 2275–2285 (2004).
48. L. Bliek, H. R. Verstraete, M. Verhaegen, and S. Wahls, "Online optimization with costly and noisy measurements using random Fourier expansions," IEEE Trans. Neural Netw. Learn. Syst. **29**, 167–182 (2016).