An efficient solver for designing optimal sampling schemes

Filip Elvander\textsuperscript{*}, Johan Swärd, and Andreas Jakobsson\textsuperscript{†}
\textsuperscript{*}Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium
\textsuperscript{†}Centre for Mathematical Sciences, Lund University, Sweden

Abstract—In this short paper, we describe an efficient numerical solver for the optimal sampling problem considered in Designing Sampling Schemes for Multi-Dimensional Data \cite{1}. An implementation may be found on https://www.maths.lu.se/staff/andreas-jakobsson/publications/.

Index Terms—Optimal sampling, convex optimization.

I. PROBLEM STATEMENT

For a background to the optimal sampling problem, see \cite{1}. Consider the signal model

\[ y(t) \sim p(\cdot; t, \theta) \]

where \( p \) is a probability density function parametrized by the sampling parameter \( t \in \mathbb{R}^n \) and the parameter vector \( \theta \in \mathbb{R}^P \). Here, \( \theta \) is the parameter of interest to be estimated. Assume that we get to choose to sample \( y \) at \( K \) out of \( N \) potential samples \( t_n, n = 1, \ldots, N \). We then want to solve

\[
\begin{align*}
\min_{w \in W, \mu \in \mathbb{R}^P} & \sum_{p=1}^P \psi_p \mu_p \\
\text{subject to} & \sum_{n=1}^N w_n F_n(\theta) e_p \mu_p \succeq 0 \quad \forall p = 1, \ldots, P,
\end{align*}
\]

(1)

where

\[ W = \left\{ w \in \mathbb{R}^N \mid \sum_{n=1}^N w_n = K, w_n \in [0, 1] \right\} \]

is the set of allowed weights, indicating which \( K \) samples that are selected, \( \mu \) is a probability density function parametrized by the sampling parameter \( t \in \mathbb{R}^n \) and the parameter vector \( \theta \in \mathbb{R}^P \). Herein, we describe how to solve (1) efficiently by considering its dual problem\textsuperscript{1}

II. DUAL PROBLEM

Consider the Lagrangian relaxation of (1) according to

\[
\mathcal{L} = \sum_p \psi_p \mu_p - \sum_{p=1}^P \left\langle G_p, \left[ \sum_{n=1}^N w_n F_n(\theta) e_p \mu_p \right] \right\rangle,
\]

where \( G_p, p = 1, \ldots, P \), are dual variables, i.e., positive semi-definite matrices of dimension \( P \times P \). Let \( G_p \) be structured according to

\[ G_p = \begin{bmatrix} \bar{G}_p & \gamma_p \\ \gamma_p^T & g_p \end{bmatrix}. \]

(2)

For notational convenience, let a dual point be denoted

\[ \bar{G} = \{G_p\}_{p=1}^P \]

(3)

and define

\[ \xi_n(\bar{G}) = \left\langle F_n(\theta), \sum_{p=1}^P \bar{G}_p \right\rangle \]

(4)

Then, for any \( w \),

\[
\inf_{\mu} \mathcal{L} = \begin{cases} -\sum_{n=1}^N w_n \xi_n(\bar{G}) - 2 \sum_{p=1}^P e_p^T \gamma_p, & \text{if } g_p = \psi_p, \\ -\infty, & \text{otherwise}. \end{cases}
\]

The infimum with respect to \( w \in W \) is given by setting the \( K \) entries corresponding to the \( K \) largest values of \( \xi_n(\bar{G}) \) equal to 1 and the rest to zero. Note that the minimizing \( w \) is not necessarily unique. Specifically, if the \( K+1 \)th largest value of \( \xi_n(\bar{G}) \) is strictly smaller than the \( K \)th largest, the minimizing \( w \) is unique. Otherwise, there are infinitely many solutions.

Algorithm 1 Sub-gradient ascent.

\[ \text{Require: Initial guess } \bar{G} = \{G_p\}_{p=1}^P, \text{ step size } \alpha. \]

\[ \text{while Not converged do} \]

\[ \text{Find } w \in \arg \min_{w \in W} -\sum_{n=1}^N w_n \xi_n(\bar{G}). \]

\[ \text{for } p=1:P \text{ do} \]

\[ \mu_p \leftarrow e_p^T \left( \sum_{n=1}^N w_n F_n(\theta) \right)^{-1} e_p. \]

\[ \text{end for} \]

\[ \bar{G} \leftarrow \{g_p\}_{p=1}^P. \]

\[ \text{end while} \]

\[ \text{return } w \in \arg \min_{w \in W} -\sum_{n=1}^N w_n \xi_n(\bar{G}). \]

\[ \]

1An implementation of the solution algorithm may be found on https://www.maths.lu.se/staff/andreas-jakobsson/publications/.

1
Thus, the dual problem is

\[
\text{maximize}_{G \geq 0} \quad \inf_{w \in W} \quad - \sum_{n=1}^{N} w_n \xi_n(G) - \sum_{p=1}^{P} e_p^T \gamma_p
\]

subject to \( g_p = w_p, \quad p = 1, \ldots, P \).

Letting \( E = e_p e_p^T \), we may express the constraint as

\[
\langle G_p, E \rangle = w_p, \quad p = 1, \ldots, P.
\]

Thus, defining the family of sets parametrized by \( \phi \),

\[
K_\phi = \{ U \mid \langle U, E \rangle = \phi, \ U \geq 0 \}
\]

and letting \( \chi_\phi = \{ G \mid G_p \in K_{\psi_p}, \ p = 1, \ldots, P \} \)

we may express the dual problem as

\[
\text{maximize}_{G \in \chi_\phi} \quad q(G)
\]

where the dual objective function is

\[
q(G) = \inf_{w \in W} \quad \sum_{n=1}^{N} w_n \xi_n(G) - \sum_{p=1}^{P} e_p^T \gamma_p.
\]

We utilize the ideas from Nedic and Ozdaglar [2] in order to maximize the dual problem [6] using sub-gradient ascent. The algorithm is summarized in Algorithm 1. A short derivation of the step is presented in the following sections.

A. Sub-gradient ascent

For a dual point \( \bar{G} \in \chi_\phi \), a sub-gradient of \( q \) in [6] at \( \bar{G} \), denoted \( \nabla q(\bar{G}) \), can be decomposed in components \( \nabla q_p(\bar{G}) \), where each component is given by

\[
\nabla q_p(\bar{G}) = - \left[ \sum_{n=1}^{N} w_n F_n(\theta) e_p \right] e_p^T \mu_p
\]

where

\[
(w, \mu_p) = \arg \min_{w, \mu_p} \mathcal{L}(\mu, w, G).
\]

As noted earlier, one may retrieve a primal vector \( w \) from this set setting the entries of \( w \) corresponding to the \( K \) largest values of \( \{ \xi_n(G) \} \) to 1 and the rest to zero. Noting that any \( \mu_p \in \mathbb{R} \) is a member of the minimizing set, one may here choose

\[
\mu_p = e_p^T \left( \sum_{n=1}^{N} w_n F_n(\theta) \right)^{-1} e_p,
\]

i.e., the \( \mu_p \) minimizing the primal objective, while still retaining primal feasibility for this choice of \( w \). Then, a dual ascent method guaranteeing that the dual variable \( \bar{G} \) is feasible may be realized according to

\[
G_p \leftarrow \mathcal{P}_{K_{\psi_p}}(G_p + \alpha \nabla q_p(G))
\]

for \( p = 1, \ldots, P \), where \( \mathcal{P}_{K_{\phi}} \) denotes projection on \( K_\phi \), as defined in [5]. How to perform this projection is described in the next section.

B. Projection on PSD cone with constraint

Consider a set of \( G = \{ G_m \}_{m=1}^{M} \) of \( M \in \mathbb{N} \) symmetric matrices \( G_m \in \mathbb{R}^{P \times P} \). Let \( C \) be the set of \( P \times P \) positive semidefinite matrices. Here, we are interested in computing the projection on the set

\[
K_\phi = \{ U \mid \langle U, E \rangle = \phi, \ U \in C \}
\]

for \( \phi \in \mathbb{R}_+ \) and a symmetric matrix \( E \in C \).

Proposition 1. The projection on \( K_\phi \), denoted \( \mathcal{P}_{K_\phi} \), is given as

\[
\mathcal{P}_{K_\phi} : G \mapsto \mathcal{P}_C(G + \lambda E)
\]

where \( \mathcal{P}_C \) denotes projection on \( C \), and where \( \lambda \in \mathbb{R} \) is the unique root of the equation

\[
\langle \mathcal{P}_C(G + \lambda E), E \rangle = \phi.
\]

Proof. See appendix.

Remark 1. It may here be noted that projecting on \( C \) is simply performed by computing an eigenvalue decomposition and setting all negative eigenvalues to zero.

C. Computational complexity

It may be noted that finding \( (w, \mu) \in \arg \min_{w, \mu} \mathcal{L}(\mu, w, G) \) is linear in \( N \) and quadratic in \( P \). Performing the gradient step is linear in \( P \), whereas the projection on the dual feasible set is \( O(P^3) \). To see this, it may be noted that in practice, one may solve the equation \( \langle \mathcal{P}(G + \lambda E), E \rangle = \phi \) using interval halving techniques, where each evaluation of the right-hand side requires computing one eigenvalue decomposition. The per-iteration complexity for this scheme is thus \( O(P^3) \).
APPENDIX

Proof of Proposition 1. By definition, \( U = P_K \phi(G) \) solves

\[
\minimize_{U \in K} \frac{1}{2} \left\| U - G \right\|_F^2,
\]

where \( \left\| \cdot \right\|_F \) is the Frobenius norm. To arrive at a dual formulation, consider the Lagrangian

\[
\tilde{\mathcal{L}} = \frac{1}{2} \left\| U - G \right\|_F^2 - \lambda \left( \langle U, E \rangle - \phi \right) - \langle \Lambda, U \rangle,
\]

with dual variables \( \Lambda \in C \) and \( \lambda \in \mathbb{R} \). We may complete the square according to

\[
\frac{1}{2} \left\| U - G \right\|_F^2 - \langle \lambda E + \Lambda, U \rangle
= \frac{1}{2} \left\| U - (G + \lambda E + \Lambda) \right\|_F^2 - \frac{1}{2} \left\| G + \lambda E + \Lambda \right\|_F^2 - \frac{1}{2} \left\| G \right\|_F^2.
\]

Then, the infimum of \( \tilde{\mathcal{L}} \) with respect to \( U \) is given by

\[
\inf_U \tilde{\mathcal{L}} = -\frac{1}{2} \left\| G + \lambda E + \Lambda \right\|_F^2 + \langle G \rangle_2^2 + \phi \lambda,
\]

which is attained for \( U = G + \lambda E + \Lambda \). Consider the dual function

\[
r(\Lambda, \lambda) = -\frac{1}{2} \left\| G + \lambda E + \Lambda \right\|_F^2 + \phi \lambda.
\]

For each \( \lambda \), this is maximized with respect to \( \Lambda \in C \) by

\[
\Lambda = P_C \left( - (G + \lambda E) \right),
\]

i.e., \( \Lambda \) is constructed from the negative part of the eigendecomposition of \( G + \lambda E \), with flipped sign. Using

\[
G + \lambda E = P_C (G + \lambda E) + P_C \left( - (G + \lambda E) \right),
\]

this yields

\[
\tilde{r}(\lambda) = \sup_{\Lambda} r = -\frac{1}{2} \left\| P_C (G + \lambda E) \right\|_F^2 + \phi \lambda.
\]

This one-dimensional criterion may then be maximized with respect to \( \lambda \in \mathbb{R} \). However, as we from the analysis obtain \( U = P_C (G + \lambda E) \), we may utilize the primal feasibility condition \( \langle U, E \rangle = \phi \). Specifically, one may seek the roots of

\[
f(\lambda) = \langle P_C (G + \lambda E), E \rangle - \phi.
\]

As \( E \in C \), \( f \) is a continuous, monotone increasing function, and \( f \) thus has a unique zero. \( \square \)

REFERENCES

[1] J. Swärd, F. Elvander, and A. Jakobsson, “Designing Sampling Schemes for Multi-Dimensional Data,” Signal Processing, vol. 150, pp. 1–10, 9 2018.

[2] A. Nedic and A. Ozdaglar, “Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods,” SIAM J. Optim., vol. 19, no. 4, pp. 1757–1780, 2009.