Krivov, Shapovalov and Dunbrack

"Improved prediction of protein side-chain conformations with SCWRL4"

Optimization of parameters via Stochastic Sampling of a Quadratic Form

Here we present a protocol that we developed to do a constrained optimization of an expensive function of multiple variables. Let a scalar function $f(\vec{x})$ be defined within an n-dimensional ball $\|\vec{x}\| \le R$, $\vec{x} \in \mathbb{R}^n$. Consider the following task

$$\vec{x} = \underset{\|\vec{x}\| \le R}{\arg\max} f(\vec{x})$$

Assume that the evaluation of this function takes a significant computational effort. Let *f* have a simple landscape so that within the ball concerned it can be roughly approximated by a quadratic form over $\vec{x}$:

$$f(\vec{x}) \approx Q(\vec{x}), \|\vec{x}\| < R$$

where quadratic form *Q* is taken in the form:

$$Q(\vec{x}) = \vec{x}^{\mathrm{T}} S \vec{x} + 2\left(\vec{h} \cdot \vec{x}\right) + d$$

Note that for a given $\vec{x}$ the form *Q* is a linear function over coefficients of the symmetric matrix *S*, vector $\vec{h}$ and scalar *d.* To find an approximate solution of the maximization task, the following procedure was used:

1. Generate random sample of points within the ball concerned:

$$X = \{\vec{x}_i\}_{i=1,N}, \quad \|\vec{x}\| \le R$$

The quality of the sample is important. In order to avoid redundant evaluation of an expensive objective function, the points should not lie close to each other. Also, for our application they should not lie too close to the bound of the ball. We place one point at the origin and for the others demand that $0.5R < \|\vec{x}_i\| < 0.76R$.

2. Evaluate the objective function, which is the side-chain prediction accuracy (average absolute accuracy) over the training set, at these points:

$$f_i = f\left(\vec{x}_i\right), \quad i = 1,2,...,N$$

3. From the linear regression, estimate the coefficients of the matrix *S*, the vector $\vec{h}$, and their covariance, $K_{S,\vec{h}}$:

$$\left\{Q(\vec{x}) \sim f_i\right\}_{i=1,N} \to S,\vec{h},K_{S,\vec{h}}$$

This is done by the classical least squares method under an assumption of the normal distribution of the underlying values.

4. Using these estimates, generate a random sample of symmetric matrices around the estimated average and covariance matrices of the dependent distribution of vector $\vec{h}$:

$$S,\vec{h},K_{S,\vec{h}} \to \left\{S_j, K_{\vec{h}}^j\right\}_{j=1,p}$$

Since we accepted that the original distribution is normal the distribution of vector $\vec{h}$ is also normal. Note that its actual characteristics depend on a particular *S*$_i$.

5. For every matrix *S*$_j$, use the corresponding covariance matrix $K_{\vec{h}}^j$ to generate a random sample of vectors:

$$S_j, K_{\vec{h}}^j \to \left\{\vec{h}_{jk}\right\}_{k=1,q}$$

The size of this sample depends on the probability density near matrix S_i. For matrices that are located in the more populated area the size of the dependent sample is larger than for those matrices that yield small values of the probability density.

6. For every matrix S_j and every corresponding vector $\vec{h}_{jk}$ consider the constrained maximization of the quadratic form:

$$\arg\max_{\|\vec{x}\|\leq R}\left\{\vec{x}^{\mathrm{T}}S_j\vec{x}+2\left(\vec{h}_{jk}\cdot\vec{x}\right)\right\}$$

This task is solved using standard Lagrange multiplier method as described in the next section. Thus we obtain a sample of points each of which represent a solution of the task that is close to the original task.

7. All solutions from the vectors $\vec{h}_{jk}$ are aggregated to estimate the average $\vec{x}^*$ and its covariance W. This average value belongs to the original ball because the last one is a convex area.

8. Finally the objective function is evaluated at the resolved average optimum location and this value is verified to be the actual maximum among all the previously sampled values. With the following definition,

$$f_{max}=\max_{\vec{x}_i\in X}f\left(\vec{x}_i\right)$$

we consider two possible cases, which are treated separately:

A. If $f\left(x^*\right)\geq f_{max}-\varepsilon$ then the resolved average optimum x^* becomes the solution of the task.

B. Otherwise consider non-empty set of “semi-optimal” points:

$$X_\varepsilon=\left\{x_i\in X:f\left(x_i\right)>f_{max}-\varepsilon\right\}$$

The value of the objective function at any of these points is close to the maximum sampled. The selection of one of them as the solution is based on its closeness to the resolved average optimum with respect to the covariance matrix:

$$\arg\min_{x_i \in X} \left(\vec{x}_i - \vec{x}^*\right)^{\mathrm{T}} \Omega^{-1} \left(\vec{x}_i - \vec{x}^*\right)$$

If the covariance matrix is singular then regularization is applied – all zero eigenvalues are substituted by small coefficient times the minimal positive eigenvalue of *W*.

We derive the actual value of threshold ε from the variance of sampled values:

$$\varepsilon = 0.01 \times \underset{\vec{x}_i \in X}{RMSD}\, f\left(\vec{x}_i\right)$$

Constrained maximization of a quadratic form

Here we describe how to perform a constrained maximization of a quadratic form required at the fifth step of the protocol. Let the linear vector space V_n be given a quadratic form:

$$F\left(\vec{x}\right) = \vec{x}^{\mathrm{T}} S \vec{x} + 2\left(\vec{h} \cdot \vec{x}\right) + c$$

Consider the following task:

$$\arg\max_{\|\vec{x}\| \le R} F\left(\vec{x}\right)$$

Since matrix *S* is symmetric we can decompose it as:

$$S = D \Lambda D^{\mathrm{T}}$$

where matrix *D* is orthogonal (containing the eigenvectors of *S*) and matrix Λ is diagonal (real eigenvalues l_k).

If $\vec{h} = \vec{0}$, then the maximization task is equivalent to

$$\arg\max_{\|\vec{x}\| \le R} \vec{x}^{\mathrm{T}} S \vec{x}$$

In this case the solution of the task is a set of normalized eigenvectors of matrix *S* that correspond to its maximum eigenvalue:

$$\left\{\vec{x} : S\vec{x} = \lambda_{max}\vec{x}, \quad w\,\|\vec{x}\| = R\right\}$$

If $\vec{h} \neq 0$, the gradient of *F* is

$$\nabla F = 2\left(S\vec{x} + \vec{h}\right)$$

It turns to zero if

$$\vec{x}_0 = -S^{-1}\vec{h}$$

If matrix *S* is negatively defined and $\|\vec{x}\| \leq R$ then $\vec{x}_0$ is the unique solution of the task.

Otherwise we use the method of Lagrange multipliers:

$$L = \frac{1}{2}\left[F(\vec{x}) - u\left(\|\vec{x}\|^2 - R^2\right)\right]$$

Its gradient is:

$$\nabla_{\vec{x}} L = \vec{h} - (u \cdot I - S)\vec{x}$$

It turns into zero if

$$\vec{x} = (u \cdot I - S)^{-1}\,\vec{h}$$

We can rewrite it as:

$$\vec{x} = D(u \cdot I - \Lambda)^{-1}\,\vec{y}$$

where we introduced a vector$\vec{y} = D^{\mathrm{T}}\vec{h}$. Substituting this into the constraint $\vec{x}^{\mathrm{T}}\vec{x} = R^2$ we obtain (after some simplifications) an explicit equation for Lagrange variable ***u***:

$$\sum_{k=1,n} \frac{y_k^2}{\left(u - \lambda_k\right)^2} = R^2$$

This equation can have up to 2*n* distinct real roots. Therefore, we obtain a set of vectors that can be enumerated to identify which of its elements maximize the quadratic form. One or several of these vectors is the solution of the task.